January 2010

Structural and Functional Studies on BK(Ca) Channels

Akansha Saxena
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

School of Engineering and Applied Science

Department of Biomedical Engineering

Thesis Examination Committee:
David Sept, Chair
Rohit V. Pappu, co-chair
Jianmin Cui
Igor R. Efimov
Garland R. Marshall
Paul Schlesinger

STRUCTURAL AND FUNCTIONAL STUDIES ON BK_{CA} CHANNELS

by

Akansha Saxena

A dissertation presented to the Graduate School of Arts and Sciences
of Washington University in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2010
Saint Louis, Missouri

ABSTRACT OF THE DISSERTATION

Structural and Functional Studies on BK_{Ca} channels

by

Akansha Saxena

Doctor of Philosophy in Biomedical Engineering

Washington University in St. Louis, 2010

Research Advisor: Professor David Sept

The long term goal of this research is to study the structure and function of the BK_{Ca} channels, by focusing on the effect of a single residue mutation, the epilepsy mutation. BK_{Ca} channels are potassium channels, activated by voltage, Ca^{2+} and Mg^{2+} ions. These factors control the opening and closing of the channel pore and thus regulate the large K^+ current passing through them. Recently a mutation, D434G in humans, was found to make the channel hyperactive and more sensitive to the Ca^{2+} ions. The single residue mutation, resulting from a substitution of an Asp to Gly, was found to be linked with epilepsy and paroxysmal dyskinesia. The central focus of this thesis is to identify the molecular mechanism behind the structural and functional changes caused by this mutation.

Using comparative modeling and molecular dynamics simulations, it is revealed that the epilepsy mutation reduces the flexibility of the channel protein and drives it to a rigid conformation. The loss in dynamics is seen around the Ca^{2+} binding site which reflects its direct impact on the Ca^{2+} activation of the channel. Comparison with

experimental results show that the change in dynamics is targeted to regions which possibly connect the Ca^{2+}–binding site to the pore and thus transfer this effect to the pore.

The thesis also presents a new method of representation of cations in computational techniques, the multisite cation model. The model presents improvement in the reproduction of accurate structural and thermodynamical properties of ion–mediated mechanisms. The successful implementation of the model in protein and water systems show that the model will prove very useful in increasing the accuracy and precision of metal mediated simulations and energy calculations.

Acknowledgments

Thank you Dave, for your immense support and motivation which has kept me going through this long and hard journey. Your guidance, technical expertise and encouragement has taught me to see and appreciate science and research. Under your guidance I have not only grown as a researcher but have also learnt important aspects of life. Your mentorship will always be reflected in all aspects of my personality and career.

A special word of thanks to my lab-mate and also my best friend Diana, who has been with me through all the ups and downs of life. Thank you for your friendship and your unconditional support through every phase of this journey.

My special thanks to Meghna and Dick Wu, for their friendship and boundless support while sharing this journey with me. Thanks to my friends, Robinson and David at University of Michigan whose friendship gave me invaluable support at at time when I needed the most. My other lab-mates, Hoon, Karthik and Frank for keeping up with my long, ever lasting discussions and providing me technical expertise.

I am very grateful to the members of the Cui lab, especially Alex Yang, Gayathri Krishnamoorthy and Dr Jianmin Cui for providing relevant insights into this investigation.

My greatest support, my grandmother and parents, who have been my constant source of inspiration. The values of hard work, commitment and success they have taught me, has always shown me the direction towards my dreams. To them I dedicate this work.

Finally, I would like to thank my sister Amisha, her husband Anurag and son Sanidhya, and my cousin Abhishek and his wife Ritu, who have been an integral part of my life while making this dream possible. Their patience, love and constant support have sustained me through difficult times and has been a source of inspiration and courage.

Akansha Saxena

Washington University in Saint Louis
December 2010

This dissertation is dedicated to my grandmother

Subhadra Saxena

and

my parents

Vandana and Anil Saxena

Contents

List of Tables

List of Figures

Preface

The work in this dissertation is divided into two major sections. The first part deals with studying the structural and functional properties of a protein, called BK_{Ca} channel. The second part talks about a different project called the multisite cation model, a new computational model of cations. Though the second part does not directly tie in with topic of the thesis, its motivation came from the BK_{Ca} channel project.

For the entire work, computational techniques were utilized and while using these tools on the BK_{Ca} channel protein, I stumbled upon an important topic in the computational methodology. I was working on demonstrating the ion mediated mechanism in the BK_{Ca} channel protein using computational tools and this required correct representation of the protein and the metal ion in the calculations. As I delved deeper into the methodology I found that several approximations were being used for these calculations. Thus, I focussed my research towards developing new methods which could improve the calculations of ion mediated molecular mechanisms and this led to the multisite cation model.

Chapter 1

Introduction

Until recently, the principal way of learning about protein functions was to gain knowledge from biochemical, genetic or structural experiments. Over the past decade, however, with the advancement of new computational technologies such as homology methods and molecular mechanics approach, our knowledge of the proteins has considerably expanded and now in addition to the structural and functional information, we can obtain the kinetic and thermodynamics data of the systems. The techniques allow the researchers to quantify the mechanisms and get results at a precision and at a time scale which was otherwise inaccessible. These techniques are particularly useful when the requirements for the expression of proteins are very demanding.

A special class of proteins, called ion channels, are complex proteins which pose several challenges in the synthesis of qualitatively and quantitatively acceptable yields of recombinant proteins. This is why these proteins represent less than 1-2% of the available crystal structures. This is in high contrast to the importance of these proteins, as they form an integral part in shaping the electric signal of of the body. Most of the cardiac and neurological diseases involve a faulty membrane proteins, and they form around 60-70% of the current drug targets. In recent years, researchers have started using computational tools to study these complex protein.

Computational techniques are, in fact, advantageous for these systems because they don't require protein expressions and they can utilize the similarities between one protein and its cousins to understand the physiological function. Membrane proteins are presumed to be descending from the same common ancestral protein and are less diverse. They have similar transmembrane helical arrangement and similar functions and hence the homology methods can be easily extended from one to the other.

In this thesis, we have utilized these characteristics of the membrane proteins and employed computational techniques to gain insight into the structural and functional relationship of an ion channel, called the BK_{Ca} channel. This thesis is divided into two projects. The first project is focussed on the BK_{Ca} channels and the second project is aimed at developing a new model for representation of cations in computational methods.

1.1 BK_{Ca} Channels

BK_{Ca} channels belong to a specialized class of membrane proteins, called ion channels. Ion channels are responsible for allowing ions to pass through the membrane, which results in generation of electric signals in our body. BK_{Ca} channels are selective to potassium ions and their activity is triggered by depolarization of the membrane. They have exceptionally high conductance (as high as 250-300 pS) and hence have been named BK_{Ca} or Big(B) Potassium (K) channels (Toro et al., 1998).

In addition to voltage, two other factors are known to enhance the opening of these channels: increase in intracellular calcium and magnesium ions. Under physiological conditions the high conductance is achieved when the depolarization occurs and also the calcium is present in high concentration. This is why these channels have been found to be in close proximity to the calcium entry points, such as the calcium channels in neurons or ryanodine receptors in smooth endoplasmic reticulum. The calcium binding sites in the BK_{Ca} channel take advantage of the high Ca^{2+} concentration in these region and help in regulating the Ca^{2+} content inside the cells. This property makes them instrumental in functioning of excitable cells and several other physiological processes of the body such as vasoregulation, neurotransmitter release, cochlear hair cell tuning, immunity, etc. Mutations in this channel have been linked to several neuronal diseases such as epilepsy, schizoprenia, etc.

1.2 Role of BK_{Ca} Channels in Physiology

1.2.1 Physiological Functions

BK_{Ca} channels are found in a wide range of cells in the human body and are involved in many physiological functions. They are particularly abundant in the smooth muscle cells and have been found to play a key role in setting the contractile tone. In neurons, they control the interspike intervals (Faber and Sah, 2003) and modulate the neurotransmitter release. Other functions include endocrine secretion, neuronal excitability, hair cell tuning, regulation of endocrine secretion, etc (Kaczorowski et al., 1996). Their role in the smooth muscles have been carefully examined and it has been shown that the BK_{Ca} channels act as negative feedback regulators of the calcium ion concentration inside the cells. We discuss this mechanism here, briefly.

Since smooth muscle cells line the blood vessels, BK_{Ca} channels find an important role in vasoregulation (Nelson et al., 1995; Standen, 2000). Calcium ion in the smooth muscle cells provides the signal for contraction. The ion, which is usually present in the plasma membrane, enters the cell through the voltage dependent L–type calcium channels. This entry increases the calcium content inside the cell and triggers more calcium ion release from the intracellular stores such as sarcoplasmic reticulum, a process mediated by ryanodine receptors. This leads to a global rise in calcium concentration by hundreds of nanomolar leading to excess of one micromolar and is sufficient to initiate contraction.For the relaxation to occur the concentration needs to be regulated and that is achieved by activation of the proximally located membrane bound calcium-activated BK_{Ca} channels (Figure 1.1). Upon activation, the high conductance of these channels elicit a potassium efflux that hyperpolarizes the membrane. This causes a decrease in the entry of calcium ions through the calcium channels as those channels are voltage sensitive and they get deactivated by hyperpolarization. The reduction in the calcium content leads to relaxation of smooth muscle cells. Thus, BK_{Ca} channels act as the negative feedback regulator in maintaining the dynamic equilibrium between constriction and relaxation of smooth muscle which is necessary for a normal vascular tone (Perez et al., 1999; ZhuGe et al., 2000; Lohn et al., 2001).

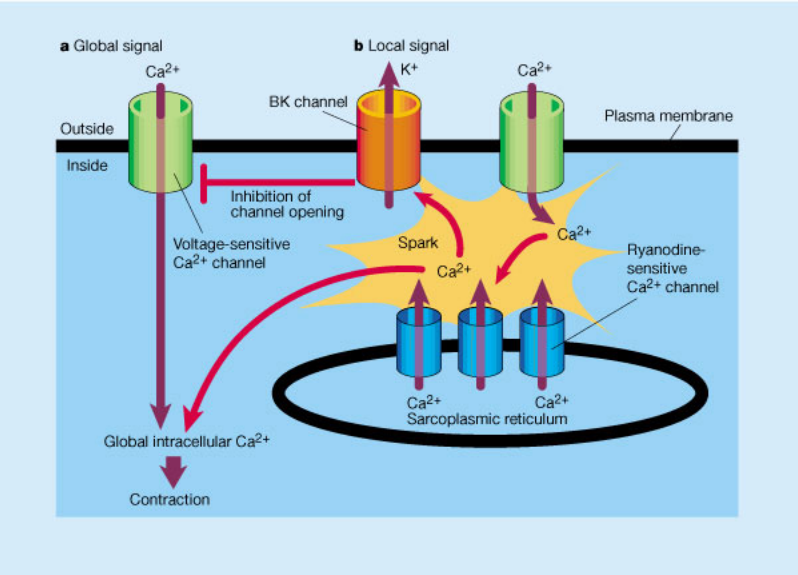


Figure 1.1: Vasoregulation by BK_{Ca} channel, Reprinted by permission from Macmillan Publishers Ltd:Nature (Standen, 2000), copyright (2000).

1.2.2 Implication in Neurological Disorders

BK_{Ca} have been implicated in various neurological diseases such as epilepsy, schizoprenia, autism and metal retardation (Du et al., 2005; Zhang et al., 2006; Laumonnier et al., 2006). Genome wide linkage scans of 13 individuals suffering with epilepsy and paroxysmal dyskinesia discovered a mutation in the BK_{Ca} channel gene which was linked with the diseases. This was a missense mutation from the base T to C in the gene which resulted in the substitution of an Asp to a Gly in the subsequent protein. Single channel studies on the mutant channels made them hyperactive and more sensitive to calcium ions (Du et al., 2005). The functional effects of this mutation have been utilized in this thesis and compared with the results obtained from the simulations. The structural and functional effects of this mutation will be discussed in detail in Section 1.6.

Another disorder which is related to these channels is schizophrenia. Antipsychotic drugs used for schizophrenia were seen to affect the BK_{Ca} channel functions and its expressions. Human postmortem brain study on patients with schizophrenia found that these patients had a lower expression of BK_{Ca} mRNA levels from the control

group (Zhang et al., 2006). Further, BK_{Ca} gene was found to be associated with autism and mental retardation (Laumonnier et al., 2006).

1.3 Structure of the Channel

1.3.1 Structural domains

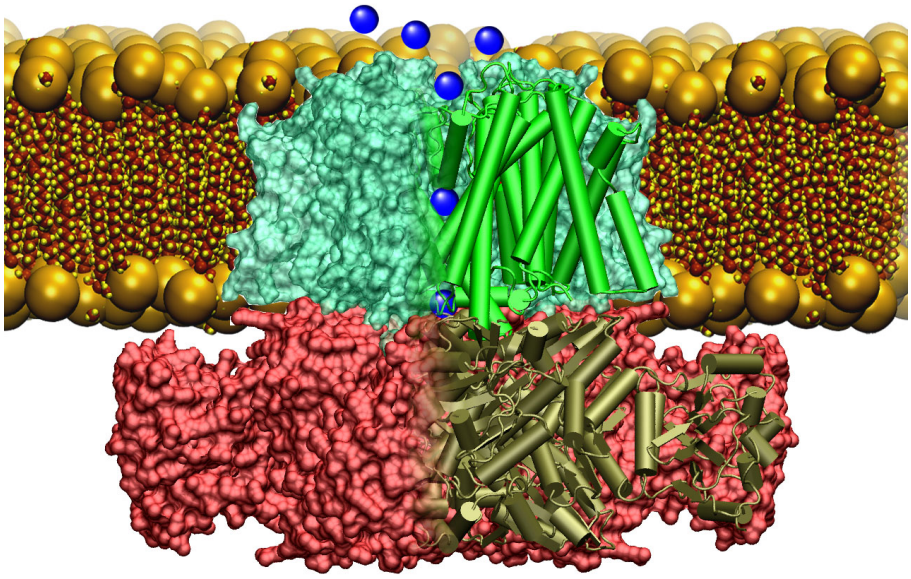


Figure 1.2: Representation of the channel structure embedded in the lipid bilayer. The channel is divided into two domains, the membrane spanning (green) and the cytoplasmic. The membrane spanning domain consists of transmembrane helices S0 to S6. The cytoplasmic domain contains the high-affinity ion binding sites.

The BK_{Ca} channel exists as a tetramer composed of four identical α-subunits (Figure 1.2). These α subunits, which form the potassium pore, were first cloned and characterized from the slowpoke locus of Drosophila (Atkinson et al., 1991); hence, the gene was named as *Slo1*. The mutation of this subunit in the slowpoke flies resulted in significant lengthening of muscle action potential, indicating that the BK_{Ca} channels played a major role in repolarization (Elkins et al., 1986).

The α subunit can be divided into two domains; the membrane spanning domain and the cytoplasmic domain. The membrane-spanning domain consists of seven putative transmembrane spanning helical segments starting from S0 to S6 (Figure 1.3). The S4 helix contains several charged residues and hence serves as the voltage sensor in the voltage activation of the channel. The segment between S5 and S6 contains the selectivity filter for the potassium ion and these three (S5, selectivity filter and S6) form the pore of the channel. The structure and arrangement of the membrane spanning domain is very similar to the voltage sensitive Kv channels.

The characteristic feature of these channels is the bulky cytoplasmic domain. Until recently, there was no structural image of the channel protein. Therefore, studies had been using *MthK* channel and *E. coli* channel crystal structure as reference for understanding the 3-dimensional view of the cytoplasmic region of the BK_{Ca} channel. *MthK* channel is a Ca^{2+}-activated *archeon* channel which has more than 40% homology with the cytoplasmic region of BK_{Ca} channel and its crystal structure was solved in 2002 (Jiang et al., 2002a). *E. coli* calcium channel also has a high homology with BK_{Ca} channel sequence and its crystal structure was solved a year earlier by the same group (Jiang et al., 2001). Based on the homology with these prokaryotic channels, it is believed that the BK_{Ca} channel gating ring consists of two RCK domains, RCK1 and RCK2 (regulators (R) of the conductance (C) of potassium (K) channels). The Ca^{2+} and Mg^{2+} activation of the channel is believed to be present in these RCK domains which host the high affinity binding sites. The structure and location of the binding sites differ considerably from the prokaryotic counterpart and are therefore, is still an ongoing area of research.

A low resolution cryogenic electron microscopy (cryo–EM) structure was solved last year which gave evidence on the arrangement of the different subunits and of the RCK domains with respect to each other (Wang and Sigworth, 2009). Very recently, two crystal structures of the channel were solved: one, in the presence of calcium ions and other without it. These structures provide the first experimentally obtained 3-dimensional view of the BK_{Ca} channel gating ring.

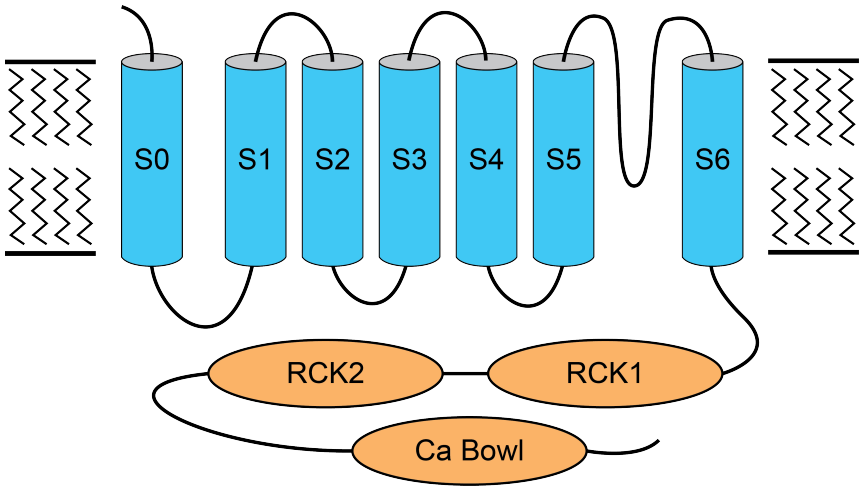


Figure 1.3: A single subunit of the channel showing the secondary structural elements. Helices S0 to S6 are part of the membrane spanning region and the remaining are part of the cytoplasmic region. The helices S0 to S4 contain the voltage sensing mechanism and S5 to S6 forms the pore.

1.3.2 Crystal Structures

A 3 Å monomeric crystal structure of the human BK channel cytoplasmic domain along with a tetrameric gating ring model of the channel based on the 6 Å structure of the Na^+ activated chicken *Slo2* channel was published this year (Yuan et al., 2010) (Figure 1.4). This structure was obtained in the presence of 50 mM Ca^{2+} ions and the the authors speculated that it represented the open conformation of the channel. The ligand was seen to bind in the RCK2 domain in a region called, the calcium bowl. Another structure, in the Ca^{2+} free form was released later (Wu et al., 2010) which represented the closed conformation of the channel.

Both the structures showed that the channel was a tetramer and each subunit contained two tandem RCK domains. These results matched strikingly well with the predictions made earlier using the *MthK* channel and the *E. coli* channel structures. Further, these structures showed that each RCK domain could be divided into three sub–domains: a N-terminal Rossmann–folded subdomain (from βA to βF), a helix cross–over domain (αF–turn– αG) and a C–terminal domain and extensive interaction

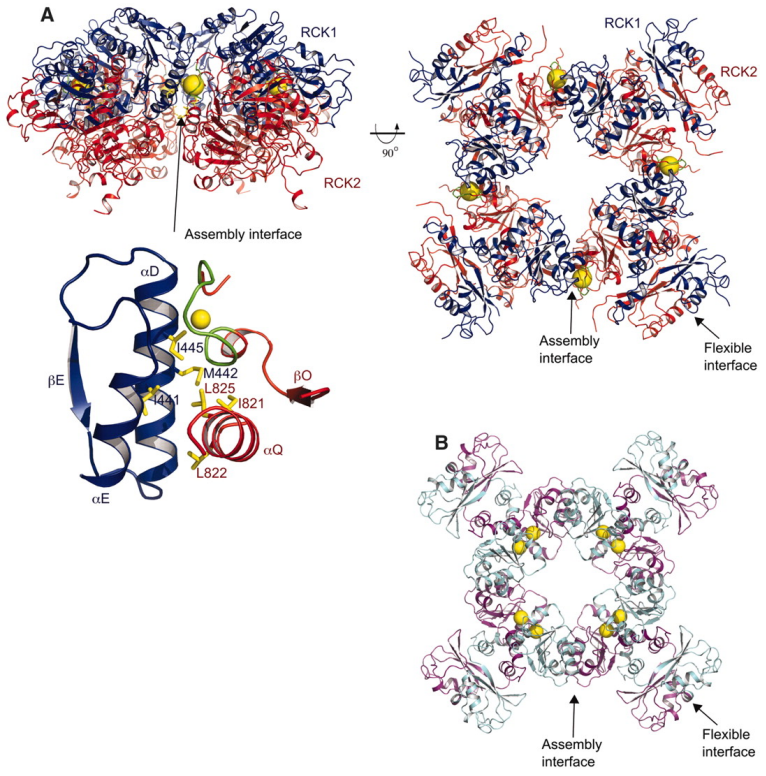


Figure 1.4: Crystal structure of Calcium bound conformation of the BK_{Ca} channel gating ring. Each subunit contains 2 RCK domains, shown in blue (RCK1) and red (RCK2). The RCK1-RCK2 domains are folded tightly against each other. The calcium ion (yellow) is seen to coordinate with the calcium bowl residues. From (Yuan et al., 2010). Reprinted with permission from AAAS.

between the two domains was observed. Both the domains formed a bi–lobed architecture and were seen to tightly fold against each other around the helix cross–over domain and the C–terminal domain.

1.4 Calcium Binding Sites

The channel contains two high affinity calcium binding sites: the calcium bowl, in the RCK2 domain (Bao et al., 2002) and Asp367 (mSlo) in the RCK1 domain (Shi et al., 2002; Xia et al., 2002). Both these sites are distal from the pore and physiological experiments have revealed that the signals from these sites are transmitted to the pore through conformational changes in the protein, which begin at the binding site and terminate at the pore (Figure 1.6). The Ca^{2+} bound crystal structure, released

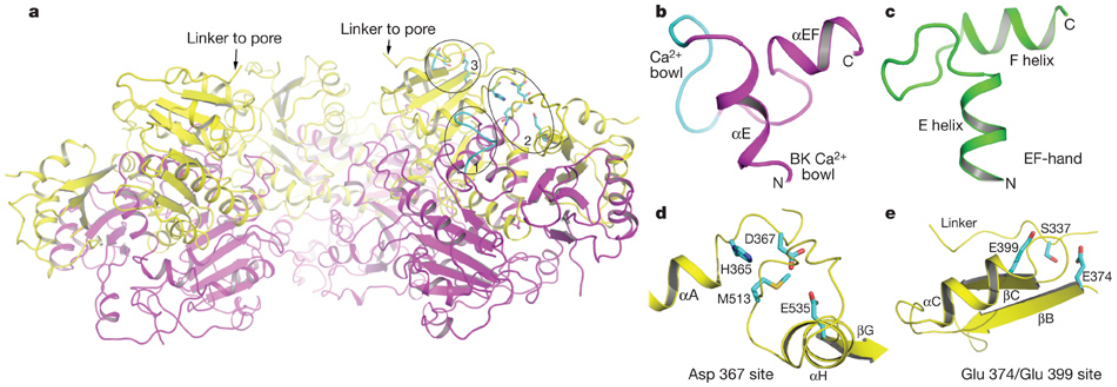


Figure 1.5: Crystal structure of the calcium free conformation of the BK_{Ca} channel gating ring. a, A side view of the BK gating ring with the top surface facing the membrane. The front subunit is removed for clarity. The circled positions represent the three Ca^{2+} binding sites on one subunit, and are labelled 1 for the Ca^{2+} bowl, 2 for the Asp367 site and 3 for the Glu374/Glu399 site. RCK1 and RCK2 are coloured yellow and magenta, respectively. d, e, Structural details of the Asp367 and Glu374/Glu399 Ca^{2+} binding sites, respectively. Reprinted by permission from Macmillan Publishers Ltd:Nature (Wu et al., 2010), copyright (2010).

recently, displayed the Ca^{2+} ion coordination the calcium bowl. It is shown to be located between the last two β strands of the RCK2 domain, the βO and βP with the oxygens from Gln889, Asp892, Asp895 and Asp897 providing the coordination. This crystal structure did not show ion binding at the other binding site, close to residue Asp367. The Ca^{2+} free crystal structure however, hints at the other binding site and they predict it might be located in the groove between the N– and C–terminal subdomains of RCK1 domain (Figure 1.5). The residues from the loop between αA and βB, the loop between αG and αH and the βG strand appear to enclose a binding region. Mutagenesis studies have also given strong evidence that several residues in this may be involved in calcium activation. These include Asp367 (Xia et al., 2002) and E535 (Zhang et al., 2010), whose mutation to Ala is seen to reduce the calcium sensitivity by ∼50-60%.

1.5 AC Region

The N-terminus of the RCK1 domain, called the AC region, was found to play an important role in imparting the calcium sensitive phenotype to the channel (Krishnamoorthy et al., 2005). Chimeras formed by swapping the AC region between two

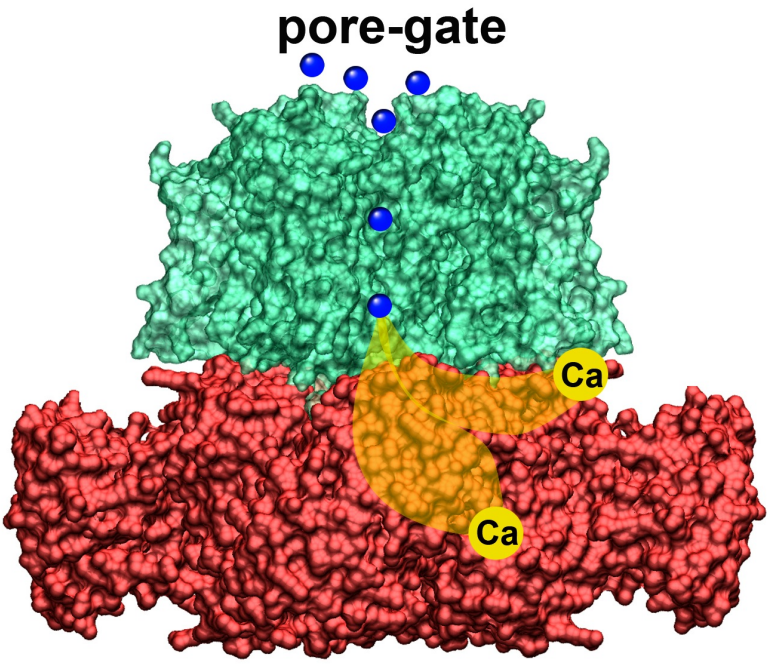


Figure 1.6: Two high affinity calcium binding sites are present in the channel structure. Here we show the probable position of these sites with respect to the full channel structure.

BK_{Ca} channel homologs, mouse and drosophila ended up transferring their respective calcium sensitive phenotype to the other. Based on the homology of the BK_{Ca} channel structure with the *MthK* channel it has become clear that this 80 residue segment forms a part of the Rossman fold structure of the RCK1 domain. It spans from the secondary structural elements βA to αC.

Within this AC region, an interloop between αA and βB, called the DRDD loop, was found to be conserved in all BK_{Ca} channel homologs (Du et al., 2005). This loop is thought to play an important role in the calcium activation of the channel. This is because the first residue in this motif, Asp367 (in mSlo1) is the putative Ca^{2+} coordinating residue (mentioned in the previous section). The third residue, Asp369 when mutated to Gly was found to make the channel hyperactive and it also increased the Ca^{2+} sensitivity of the channel. This mutation has been linked with epilepsy and paroxysmal dyskinesia in humans (Section 1.2.2). The structural location, chimera experiments and the presence of key residues makes AC region the focus of structural and functional studies on these channels. Several studies have successfully targeted

at the AC region to provide meaningful insights into the channel function (Shi et al., 2002; Krishnamoorthy et al., 2005; Fodor and Aldrich, 2006).

1.6 D369G Mutation

Studies on the BK_{Ca} channel discovered a critical mutation, D369G (D434G human; D369G, mouse), which significantly altered the function of the channel. Genetic linkage studies on human subjects showed that this mutation was linked with the human syndrome of coexistent generalized epilepsy and paroxysmal dyskinesia. The mutation involved a heterozygous transition from A to C in the BK_{Ca} channel gene KCNMA1, producing a Gly instead of an Asp at the corresponding amino acid location. Single channel recordings in heterologous expression systems demonstrated that the mutation resulted in increased channel activation, higher steady state open probabilities and an increased Ca^{2+} sensitivity (Figure 1.7) as observed from the left shift of the Ca^{2+} dependent G-V curves (Du et al., 2005; Yang et al., 2010). Mice knockout studies performed earlier have revealed that an increased activation of the BK_{Ca} channels can sharpen the action potential and facilitate high frequency firing in neurons (Brenner et al., 2005). However, the relation between the enhanced activity of the BK_{Ca} channels and the occurrence of epilepsy is an ongoing area of research.

Investigation of the molecular mechanism underlying the gain-of-function effect of the mutation revealed that the channel experienced an increase in the gating event. The gating of BK_{Ca} channels can be determined by three equilibria, as described by the Horrigan-Aldrich allosteric (HA) gating model (Horrigan and Aldrich, 2002): a central "closed-to-open" step (also called intrinsic gating), voltage sensor activation and Ca^{2+} binding. Kinetic analysis in the context of this model revealed that the mutation brought the following changes: two fold increase in the intrinsic gating (from $6.6e^{-7}$ to $1.65e^{-6}$) and two fold decrease in the Ca^{2+}-dissociation constants of both closed and open states (Wang et al., 2009). Increase in the closed to open equilibrium constant and an increased open probability of the channel opening seen in the single channel recordings, indicated that the mutation could drive the channel protein towards an open conformation. The decrease in the Ca^{2+}-dissociation constants of both the closed

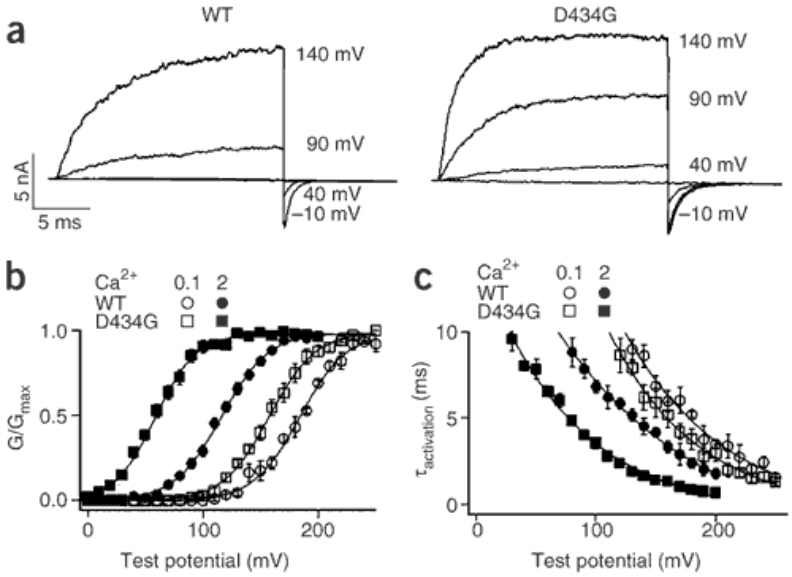


Figure 1.7: Single channel recording of the epilepsy mutation. Current traces of the wildtype (left) and D434G (right) channels at Ca^{2+} concentration of 2 μ M. Mean G-V relations of wildtype (WT) and D434G mutant channels at Ca^{2+} concentrations of 0.1 and 2 M. Reprinted by permission from Macmillan Publishers Ltd:Nature ((Du et al., 2005)), copyright (2004).

and open states and an increase in the calcium sensitivity of the channel activation showed that the mutation was targeting the Ca^{2+} activation pathway.

We talk about this mutation here because the studies mentioned above, hint at an important molecular mechanism, which is probably being disrupted by the mutation. The mutation is located in the putative Ca^{2+}-binding domain, just 2 residues downstream of Asp367, the Ca^{2+} binding partner; it is part of the DRDD loop, which is conserved in all BK_{Ca} channel homologs; and it renders the channel more sensitive to the calcium ion. Focusing on the changes produced by this mutation can give insights into the calcium activation mechanism of the channel. My thesis is aimed at inspecting the changes produced by this mutation and understanding the calcium activation mechanism in the course of the study.

1.7 Synopsis of Thesis

1.7.1 BK_{Ca} Channel Projects

The focus of this thesis is on understanding the calcium activation mechanism of the BK_{Ca} channels by probing the role of the important mutation, D369G or the epilepsy mutation in changing the function of the channel. The research has been done by using computational techniques such as homology methods and molecular dynamics simulations. The crystal structure of these channels was not known for a very long time, so the computational methods have been advantageously used in this thesis to obtain important insights into the structure and function of the channels. The study is divided into parts: the **first part** is focussed on the AC region, an important protein segment in the N–terminus of the RCK1 domain and the **second part** is expanded to the entire subunit. Computational techniques have been applied to predict the structure of the AC region in the first part since the crystal structure was unknown at that time. The structure and dynamics of the resulting models are then compared between the mutant and the wildtype protein to identify the effect the mutation. These results are then compared with the findings from the Cui lab, who performed electrophsyiological experiments on the AC region and reported the functional comparison of the mutation.

The structural and dynamical comparison between the mutant and the wildtype protein in the second part have utilized the X-ray crystallographic data instead of the homology model since two crystal structures of the BK_{Ca} channel were released recently. The results from this part give a detailed explanation of the effect of the mutation on the entire subunit of the channel and along the way give an understanding on the channels function.

1.7.2 Multisite Cation Model

The last part of the thesis answers a different question where the research is focussed on developing a new computational model of the cations which can produce accurate values of the ion interactions with protein and other biomolecules. The motivation for

this project came from a computational study involving prediction of Ca^{2+}-binding site in the BK_{Ca} channel protein. The study revealed several limitations in the calculation of interactions of ion with its environment.

Ions play a fundamental role in biological functions and their interactions with proteins could be responsible for a host of important mechanisms such as modulating protein stability, triggering a conformational change or regulating catalytic activity of enzymes. Probing these roles require highly accurate computational models of ions which can reproduce physically realistic descriptions of these mechanisms. In the past few decades several fixed-charged non-polarizable models have been developed but obtaining a balanced description of ion-ion, ion-water and ion-protein interactions still presents a difficult challenge. Usage of *abinitio* quantum mechanical approaches or polarizable force fields have been successful to some extent, but become computationally expensive. The work in this thesis proposes a non-polarizable variable-charge ion model which does not require sophisticated calculations but can still produce physically realistic values. The model has been tested in simulations with water and proteins and it is seen to show promising results in accurately reproducing both thermodynamical and structural properties of the interacting environment. Chapter 5 gives an extensive detail on the development of this model and its applications in the different systems.

1.7.3 Specific Aims

The work in this thesis can be summarized in the following Aims:

- **Aim 1**: Does the epilepsy mutation affect the structure and dynamics of the AC region? (Chapter 3)
- **Aim 2**: Does the epilepsy mutation affect the structure and function of the entire RCK1 domain of the BK_{Ca} channel? (Chapter 4)
- **Aim 3**: How can a multisite cation model improve the accuracy of cation-protein interactions in molecular mechanics calculations? (Chapter 5)

Chapter 2

Computational Methods

2.1 Introduction

Computational techniques provide significant advantage in systems where the structural data is not easily available and they also provide a means to identify the intermediates of mechanisms. Research on ion channel proteins has been significantly impacted by the challenges associated with the purification of these proteins, which results in lack of structural data. Thus, to date only ∼25% of the identified ion channel proteins have been cloned and functionally tested. This deficit, however, is being reduced by using computational techniques as these tools allow us to exploit the similarities between proteins. Additionally, important insights into the dynamics of the channel protein domains can be gained through these techniques. The purpose of this chapter is to outline the basic methodology of the commonly used computational techniques. The first section features a discussion of biomolecular structure and homology modeling techniques, and then the focus goes to molecular dynamics and sampling of protein conformational space which has been predominantly used in this thesis. The last sections talks about the calculation of free energy by using the thermodynamic integration method.

2.2 Homology Methods

As each genome is sequenced we are faced with the daunting task of digging for useful information in this growing ocean of letters. At the time of writing this document,

GenBank (Benson et al., 2009) reports nearly 100 million sequences and the Protein Data Bank contains almost 55,000 protein structures (Berman et al., 2000). Due to this amount of information, it is next to impossible to manually sort through, order and correlate all of the data. Thanks to major improvements in computing power and algorithms, we can easily handle such large-scale data and derive useful information. Computational modeling has become an essential tool in guiding and enabling one to make rational decisions with respect to hypothesis driven biological research. In parallel, the wide availability of web-based applications has produced several computational tools as online servers, which are for the most part user friendly, and have thus made it more amenable for researchers to use. Indeed, most of the tools and protocols that are discussed in this chapter can be accessed and utilized by just using a modern day laptop with an internet connection. The purpose of this chapter is to give the readers a flavor of the different computational methods that involve protein modeling, so genomics will be skipped and the discussion will directly dive into proteomics. There are numerous tools each with its unique protocol to predict the desired property and it is beyond the scope of this chapter to discuss each tool in detail. For an exhaustive list of other tools, readers are referred to sites such as Expasy (Gasteiger et al., 2003), NCBI, EMBL-EBI, Biology WorkBench (Subramaniam, 1998) and references (Kretsinger et al., 2004; Madhusudhan et al., 2005) (see Table 1 in (Saxena et al., 2009)).

2.2.1 Sequence Analysis

It is possible to get general information about a protein function by identifying certain motifs or domains in its sequence. For example one can calculate the hydrophobicity of an amino acid at a specific position and thereby create a hydropathy plot/scale for a given sequence. Using such information, one could get an idea whether a sequence segment of a protein can be on the protein interior, protein surface, or perhaps a transmembrane segment. One of the tools TMHMM (Krogh et al., 2001) predicts the transmembrane segments by applying a Hidden Markov Model (for more details, see for example (Punta et al., 2007)). If interested in predicting possible post-translational modification sites or DNA-binding motifs or signaling sites, one can scan the query sequence against protein databases such as Prosite (Hulo et al.,

2008), Prints (Attwood et al., 2003), or InterPro (Hunter et al., 2009). PredictProtein (Rost et al., 2004) and SCRATCH (Cheng et al., 2005) are online servers that can analyze the sequence by submitting the sequence to other prediction tools that can each analyze the sequence for unique features.

Since all these tools are based on statistics, it is always recommended to try multiple software packages to look for consensus and inconsistencies in the individual predictions. Many meta-servers like InterProScan (Zdobnov and Apweiler, 2001) or PredictProtein (Rost et al., 2004) enable the user to simultaneously submit the query sequence to multiple online tools from one central website. Websites like EMBL-EBI (Andrea et al., 1983), Expasy (Gasteiger et al., 2003), NCBI as mentioned above maintain a list of tools available for various forms of analysis. At the end of the day, one has to keep in mind the limitations of these tools and that the accuracy of any tool is never 100%. As mentioned above, all these prediction algorithms are based on statistical analysis of available data across multiple organisms. If any of these tools lack the specificity for the organism of interest, the confidence in its predictions may be more questionable. One should always cross check the predictions with the available knowledge of the system and see if it is compatible with the system that is being studied.

A more common and direct approach towards prediction of protein function from sequence is through searching the sequence databases such as NCBI or Swiss-Prot (Bairoch et al., 2004), using tools such as BLAST (Altschul et al., 1990; McGinnis and Madden, 2004) or FASTA (Pearson, 1990), for closely related sequences to your query. The best-case scenario would be if a query picks out sequences (hits) from the database that are both functionally well characterized and share a high sequence identity. Based on high sequence similarity to the hit, it can be generally assumed that the query protein will have a similar fold and hence might belong to the same structural family of a given hit and, depending on the extent of sequence identity, possibly have a function similar to the hit. The above stated generalization of sequence-structure-function relationship has its root in the process of divergent evolution. In divergent evolution two proteins that share sequence similarity diverged from a common ancestor and since structure diverges slowly than sequence they should also have similar fold. In general, two sequences sharing more than 40% sequence identity share similar structural fold (Davidson, 2008). Yet this simple sequence, structure, function

relationship is always not true (Roessler et al., 2008). A famous example would be myoglobin and hemoglobin, proteins that are similar in structure and function yet have only 20% sequence similarity. On the same note convergent evolution can result in proteins with very different structural folds but similar function (e.g. subtilisin). These are exceptions to the standard rule that are to be borne in mind when analyzing two sequences and inferring function or fold on the basis of sequence similarity.

In a less than ideal scenario, a query fails to generate any significant hits or the generated hits are of low sequence similarity (<25%) to a given protein and a collection of diverse proteins that cannot be used for any further inferences of functions or folds. The reasons behind this could be many. The protein of interest and its supposed homologs may have acquired major insertions and deletions that make the alignment between the two problematic. Another possibility is that the query protein is a multi-domain protein and the domains are not properly aligned with sequences in the database. For such difficult queries, instead of doing a pairwise search as discussed above the search is done based on a pattern or a profile generated from a multiple sequence alignment (Pei, 2008) based off of the query protein sequence. ClustalW (Larkin et al., 2007) is one of the most widely used tools for generating multiple sequence alignments. Software tools such as PSI-BLAST (Altschul et al., 1997), PHI-BLAST (Zhang et al., 1998) and HMMER (Eddy, 1998) are packages that are used to perform and enable pattern/profile/blocks based searching. The patterns are generated using methods like dynamic programming, Hidden Markov Models (HMM), or genetic algorithms. Using these multiple sequence alignment based pattern or profiles increases the search space of the query in a database as it increases the combination of sequence space that can be matched in a database. These patterns/profiles are searched against protein databases that incorporate these profile/patterns in their own creation and classification of protein sequences. These include databases such as Prosite (Hulo et al., 2008), Pfam (Finn et al., 2008), SMART (Letunic et al., 2009), and InterPro (Hunter et al., 2009). It is to be noted that since these databases were built based on different protocols, the contents do differ. Hence, it is important that the query is searched against multiple databases and the user should apply a consensus approach in interpreting the results.

Finally, it is crucial to know if the alignment predicted between the query and the hit is statistically significant and it is not some random match. The statistical significance

of an alignment obtained is represented as expectation value or E-value. Basically, the E-value of an alignment having score X is the number of times one expects to find alignments having score equal to or greater than X in a comparison of random sequences similar to the query. The lower the E-value, the better the alignment that is predicted. Tools such as Jalview (Waterhouse et al., 2009), T-coffee (Notredame et al., 2000) and Chimera (Pettersen et al., 2004) can be used to visualize and edit multiple sequence alignments. Choosing the right homolog for your query is important because it forms the first step for homology modeling discussed below.

2.2.2 Secondary structure prediction

The secondary structure prediction of your query sequence can also aid in determining the structural fold of the protein. For example, if the secondary structure profile of the query matches with another family with a protein of known structure then probably the query protein have a similar structure. Use of secondary structure information can also improve the sequence alignment of the query with its homolog. To aid secondary structure prediction analysis, only primary elements like alpha-helices and beta-strands are examined while other structural features are considered to be coils. Methods such as nearest neighbor, neural networks and HMMs are used to generate these predictions. The most commonly used secondary structure prediction software include JPRED (Cole et al., 2008), PSIPRED (Bryson et al., 2005), PHD (Rost and Sander, 1993), or SSPro (Cheng et al., 2005). Some methods can also provide information on the solvent accessibility of the amino acids in the query. It should be noted that the current prediction protocols have an accuracy rate of 65-75% with slightly higher accuracy in the prediction alpha-helices. Hence, it is again important to test the predictions using multiple software packages and take a consensus approach in interpreting the results.

2.2.3 Tertiary structure prediction

Predicting the 3D-structure of protein from an amino acid sequence is the holy grail for structural modelers. Having a 3D-structure of your query protein can enhance greatly a hypothesis driven biological research. Depending on the quality of the

model produced, it can be used to study the electrostatic profiles under different conditions (Sali et al., 1993; Diraviyam et al., 2003), prediction of possible ligand or small molecule binding sites, determination of residues for site-directed mutagenesis, refining of X-ray and NMR structures, and a host of other possibilities. The methods to predict 3D-structure of a protein broadly falls into three categories: 1) homology or comparative modeling, 2) threading, and 3) *abinitio* or de novo modeling. The best method to choose for building a model depends on how evolutionarily diverged the query protein with respect to other family of proteins in the database that has a representative structure. The most straightforward case is when the query protein produces significant hits of high sequence similarity (generally >40%) and one of those hits also has a representative X-ray or NMR structure. In this case, opting for homology modeling can produce reliable models that are within 2 Å RMSD of the experimental structure. For instances of low sequence identity, for example where a multiple sequence alignment approach was required to produce reliable hits, homology modeling based on a single structure cannot be reliable. However, current homology modeling protocols can produce good structural models for such proteins for even these low sequence similarity cases by using multiple templates to model different regions. For the most difficult case of low sequence similarity, alternate model building methods such as threading and *abinitio* modeling methods can be used to produce structural models of the query.

Homology modeling

The aim of homology modeling is to model or predict the structural coordinates of a query protein based on the known structure of a sequence homolog (generally referred as the template). Since the model produced can be only as good as the template, it is imperative that rigorous analysis is done before choosing the template. This is achieved by searching the query sequence on various databases and using different search protocols, as discussed above in sequence analysis. In cases where multiple templates are available, it is usually best to use the template with the highest resolution and fewest gaps in sequence/structure. One should also be aware of the environmental conditions under which the template structure was solved. After selecting an appropriate template, the next step involves determining the best alignment between the template and query. The default alignment that is produced by

sequence search tools may not be optimal, and the use of profile/patterns derived from multiple sequence alignments could again enhance the quality of alignment for less similar sequences. As with choosing the template, threading can also help in improving the sequence alignment. Threading adds in the structural information to sequence based alignment. Programs such as FUGUE (Shi et al., 2001) and SALIGN (implemented in MODELLER (Marti-Renom et al., 2000)) use an intermediate protocol where the structural information is used in a profile based alignment. Once an optimal alignment is constructed, this information is fed into homology model building software. There are online model building servers such as SWISS-MODEL (Schwede et al., 2003) or 3D-JIGSAW (Bates et al., 2001) or stand-alone software such as MODELLER that can be used. In MODELLER, the model is built based on spatial restraints resulting from the query-template alignment. It can also perform de novo loop prediction, using multiple templates to construct a model. The final model is generated using conjugate gradient and molecular dynamics simulation (simulated annealing). In analyzing the final model one has to keep in mind that in addition to errors from an incorrect alignment or a non-ideal template, there are also errors inherent to model building method that can result in backbone distortions even in conserved regions and side-chain packing or conformation errors. In such instances protein optimization programs such as PLOP (Jacobson et al., 2002b), both of which can do multiple cycles of side-chain optimization and minimization, can be used to improve the model.

Threading

This method is typically used in cases where there is no significant homology predicted (<20%) between your query and any sequence in the database, but it can also be employed to improve the alignment of low homology query-template sequences. Briefly, the query sequence is threaded through all available folds in the structural database and a score for each fold is calculated using some suitable scoring scheme, and the fold that gives the best score is assumed to be the fold for the query sequence. For more details in the methods that are used in threading the sequence and scoring algorithms the reader is referred to (W.Mount, 2001). Some of the popular online servers for threading includes 123D (Alexandrov, 1996), pGenTHREADER (Lobley et al., 2009) and 3D-PSSM (Kelley et al., 2000). Since this method is more often used

when the sequence similarity is very low, interpreting details beyond the fold of the protein, like side chain interactions, are not reliable. In this case protein optimization software can again be useful in improving the structure.

Abinitio modeling

uses a combination of statistical analysis and physics based energy function to predict the native fold of a given sequence. *Abinitio* modeling is preferred in predicting the structure of a sequence when no suitable template is found or if it is known that the query adopts a different fold than the predicted template in spite of the sequence similarity. The various different *abinitio* algorithms use statistical information, secondary structure prediction, and fragment assembly for fold prediction. Also common to all algorithms is the simplified representation of the protein to keep the prediction problem tractable. For more details on the different *abinitio* prediction methods the reader is referred to (Hardin et al., 2002). Some of the software that are currently used for *abinitio* structure prediction include Rosetta (Rohl et al., 2004b) and I-TASSER (Zhang, 2008).

Rosetta is one of the more widely used packages. It uses a fragment based assembly protocol to predict the full structure. Here, the fragments form a reduced representation of the protein. A key assumption in Rosetta is that the distribution of structures sampled by a particular sequence fragment is reasonably represented by the distributions of conformations adopted by that fragment and closely related fragments in the protein structure database. Fragment libraries are then built based on the protein structure database. The reduced conformational search space of the query is searched using the Monte Carlo algorithm with an energy function that is defined as the Bayesian probability of sequence/structure matches (Simons et al., 1997). This produces compact structures that have optimized local and nonlocal interactions. Based on the observation that folding of small protein predominantly follows a single exponential process, conformational search is achieved by running short simulations. In general 1000 short simulations are performed independently (Shortle et al., 1998). The resulting structures are clustered and generally the central model representing the largest cluster is chosen as the best predicted structure for the query sequence.

There are important limitations of this fragment based *abinitio* prediction. The conformational search space of fragments is pre-determined and long range interactions within the protein are not included in the structure prediction of fragments.

Structure validation

It is critically important that the predicted models, produced using any of the methods discussed above, should be further examined to increase confidence in the structural model. Tools such as PROCHECK (Laskowski et al., 1993) and WHATIF (Vriend, 1990) can check multiple structural variables in the model against expected/reference values for the same. PROSA (Wiederstein and Sippl, 2007) uses knowledge based potentials of mean force to evaluate model accuracy. It also gives the Z-score for the model that indicates overall model quality. An anomalous Z-score indicates problem with the structural model. Verify 3D (Luthy et al., 1992) is another tool that utilizes the location and environmental profile of each residue in relation to a set of reference structures to predict the quality of the structure. It should be noted that performing the validation analysis on both the predicted model and template and comparing the results can be useful, since the model can only be as good as the template.

2.2.4 Conclusions

As the number of experimentally solved structures and sequences deposited into their respective databases increase in number, the prediction accuracy of the sequence analysis and structure prediction tools will continue to advance. Increased availability of cheap and fast computational resources, as well as improvement in prediction algorithms will enable more exhaustive and accurate prediction in less time. In this regard, the Critical Assessment of techniques for protein Structure Prediction (CASP), a large scale meeting that evaluates the current status of prediction algorithms, is an excellent resource for evaluating the current state of the field. At the end of this biannual meeting, the performance based rankings results are published in the official website (http://predictioncenter.org/), and this resource is perhaps the best source for up-to-date information on current methods and algorithms.

2.3 Molecular Dynamics

We have seen how we can obtain a 3-dimensional structure of a protein from predicting the positions of the nuclear centers of all atoms that make up the molecule. This structure is helpful in understanding the relative position of the atoms, but this is only a static picture and does not tell us how the proteins moves and functions. More than four decades ago, scientists demonstrated that if we are able to calculate the relative positions of the protein atoms at small intervals of time, we can then predict the behavior of the atoms over a longer time scale. The principal tool for such calculations is the method of molecular dynamics (MD) simulations that was first introduced in the late 1950s by Alder and Wainwright (Alder and Wainwright, 1959, 1960). In the 1960s, Rahman carried out the first realistic MD simulations with liquid argon and liquid water (Rahman, 1964; Rahman and Stillinger, 1974), and later in 1976, the first protein simulation was performed by (McCammon et al., 1976, 1977). Since then, researchers have used MD to investigate the structural, dynamic, and thermodynamic properties of biological molecules, including such items as the characterization of protein folding kinetics and pathways, protein structure refinement and protein-protein interactions.

The goal of this section is to give an overview of the basic methodology required for studying the protein dynamics using MD simulations. The basics of MD are discussed first which is followed by a description of the methodology required to run the simulations. Thereafter, various analysis tools are presented which can be used on the resulting trajectory to study the dynamics and flexibility of the protein.

2.3.1 Molecular mechanics

The basis of the MD technique is little more than the integration of the Newton's equation of motion to calculate the positions of the atoms over time. The behavior of the atoms in such a calculation can be likened to the movement of billiard balls (Leach, 2001). Of course billiard balls move in straight lines until they collide with each other and the collisions changes their direction, but in our case, the atoms experience a varying force in between collisions. To account for the effects of these changing

forces and to get more realistic dynamics, the equation of motion are integrated over short time steps (typically 1 to 2 fs) such that the forces can be regarded as constant. The integration yields a series of conformations in time that reveal the realistic movement of the atoms and the protein as a whole. Newtons equation is a second order differential equation (the acceleration a is the second derivative of position) and this means that we are required to provide the positions of each atom, from a crystal or model structure, and their individual velocities, calculated from a thermal distribution. The forces that act on each atom, F, are determined from a molecular mechanics potential by taking the negative gradient (i.e. $F = -\nabla U$). A discussion of molecular mechanics potentials could alone fill multiple chapters, however it is sufficient to say that these parameter sets are based on first-principles physics but are parameterized empirically through detailed comparisons with many experimental measurements.

All molecular mechanics potentials deal with two primary classes of interactions: bonded interactions and non-bonded interactions. The bonded interactions are composed of a bond stretching term for two covalently bonded atoms (see Figure 2.1), an angle bending term for three consecutively bonded atoms, and a torsional term for four consecutive atoms. Since the integration time step is small and the perturbations about the equilibrium point are typically very small in proteins, these potentials are typically modeled as a harmonic springs. The use of such effective potentials makes calculations much easier and faster. The non-bonded interactions capture longer-range interactions within the protein. The major non-bonded forces include the electrostatic interactions, based on the Coulombs law, and van der Waals interactions, usually based on a Lennard-Jones potential. The following equation sums up these interactions with the first three terms representing the bonded terms while the last two representing the non-bonded interactions: The most commonly used force fields for MD include AMBER, OPLSAA, CHARMM, AMOEBA, and GROMOS ((Ren and Ponder, 2003; Duan et al., 2003; Hess et al., 2008; Brooks et al., 2009)). A nice review on the usage of different types of force fields and their development can be found in (Ponder and Case, 2003) and (Cheatham and Young, 2000).

An important component of MD simulations is the definition of a suitable environment for the protein. For cytoplasmic proteins, this means immersing the protein in water with suitable concentrations of salt or other ions, and maintaining the proper

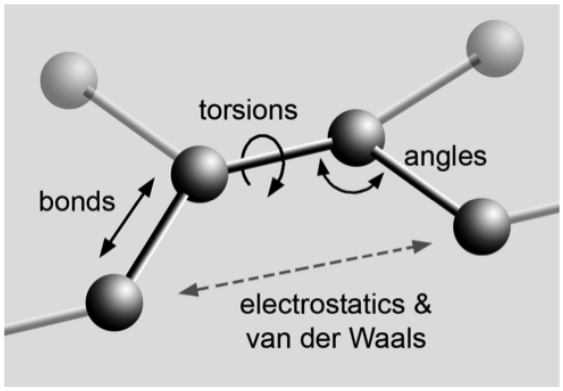


Figure 2.1: Representation of the bonded and non-bonded interactions used in molecular mechanics force fields.

temperature and pressure. There are several water models available in the literature including SPC, TIP3P, TIP4P, POL3, and AMOEBA (Berendsen et al., 1981; Caldwell and Kollman, 1995; Jorgensen et al., 1983; Jorgensen and Madura, 1985; Mahoney and Jorgensen, 2000). SPC and TIP3P are the simplest models that represent the water molecule with 3 interaction sites. The other models are more sophisticated as they include dummy atoms at the lone pair positions, which improve the dipole and quadrupole moments of the water molecule. POL3, SPC/E and AMOEBA contain additional terms to account for the polarizability. Membrane proteins are treated differently since they have to be immersed in a lipid membrane first, before placing the whole system in a water bath. Several models for the lipid membranes are available such as DPPC, DMPC, POPC, DLPE, DOPC, or DOPE (de Vries et al., 2005; Heller et al., 1993; Tieleman and Berendsen, 1996; Tieleman et al., 1997).

2.3.2 Setting up and running simulations

Many MD packages are available for no cost or with modest academic fees. Popular packages include AMBER, CHARMM, GROMACS, NAMD, and TINKER. Although there are small differences in the implementation depending on the software package being used, the simulation procedure can be divided into four basic steps: preparation, minimization, heating, and production.

Preparation: The first task of the simulation is to obtain a structure of the protein of interest. This is typically a pdb file that contains a list of all atoms in a protein and their 3D coordinates. The protein structure needs to be complete (all atoms), and missing atoms such as hydrogens can be added by using the preparatory tools of the MD software. It is also important to check the protonation states of the histidines, the existence of disulfide bonds and any potential post-translational modifications. Next, the protein is immersed in a water bath using any of the water models mentioned above. If it is a membrane protein then it is first embedded in a lipid bilayer and then the whole system is submerged in water. Salt ions can be added to more closely mimic physiological conditions.

Minimization: The starting structure almost assuredly has small atomic clashes, strained bonds and angles, or other potential problems. We need to resolve these issues before starting our simulation, and we do that by minimizing the potential energy of the system, effectively moving all bonds, angles, etc. to their equilibrium values. There is variable quality in the minimization routines for the various MD packages, but as long as major atomic clashes are removed, we should be able to proceed to the next step.

Equilibration: Since we are normally trying to connect simulation results with wet-lab experiments, we need to match the experimental conditions as closely as possible. The minimized protein structure can be viewed as being at 0 K, but we need to heat the system to a "normal" temperature of perhaps 300-310 K. Since the MD protocol is an equilibrium method, we need to slowly perturb the system, usually heating the system in 50 K steps for short periods of time (20-50 ps) until we reach our desired simulation temperature. Once we reach our production temperature, we need to allow the system to equilibrate to again remove any artifacts. The time required for equilibration is a point of debate in the simulation community and depending on the system, equilibration times may range from 100 ps to 50 ns, or more. When in doubt, more equilibration is certainly the safest route to follow.

Production: Now the system is ready to start the production run. Depending upon the computer resources, the simulations can be divided onto parallel processors to increase the speed. Just as with the equilibration step, the simulation time will depend on the size of the system and what the ultimate goal of the simulation is.

The next section will discuss some of the analysis tools that can be used to evaluate the simulation results.

2.3.3 Simulation analysis

Equilibration measures: As the simulation progresses, the protein evolves from the minimized state, attains a state of equilibrium and then begins to fluctuate around this point. One measure used by researchers is to study the evolution of the dynamics of the system is the Root Mean Square Deviation or RMSD of the protein relative to the starting structure. The RMSD is defined as the average deviation of all atoms from their starting position. The formula used for this analysis is

$$RMSD(t_1, t_2) = \sqrt{\frac{1}{M}\sum_{i=1}^{N}(m_i||r_i(t_1) - r_i(t_2)||^2)} \tag{2.1}$$

where M is the total mass and $r_i(t)$ is the position of atom i at time t. The calculation can be performed using any set of atoms, however the backbone atoms or just the α carbons are the most common choices. A typical RMSD plot is shown in Figure 2.2, in this case starting from the equilibration phase of the simulation for a protein of 140 amino acids. During the initial equilibration phase, the protein fluctuates significantly, but after about 50 ns it settles down at a steady value and could be considered to have reached equilibrium. One issue with respect to RMSD is that is depends on the reference state (in the case of Figure 2.2, the structure at the start of the production phase). For this reason, it can be a non-ideal measure for equilibrium, and many researchers use methods such as principle component analysis (discussed later).

Root mean squared fluctuations: Root mean squared fluctuation, commonly known as RMSF, is a tool which quantifies the dynamics of the polypeptide backbone by finding the extent of movement of each residue around its mean position, throughout the length of the simulation. The formula used for this analysis is

$$RMSF(r_i) = \sqrt{\sum_{i=1}^{n}(r_i(t) - \hat{r_i})^2} \tag{2.2}$$

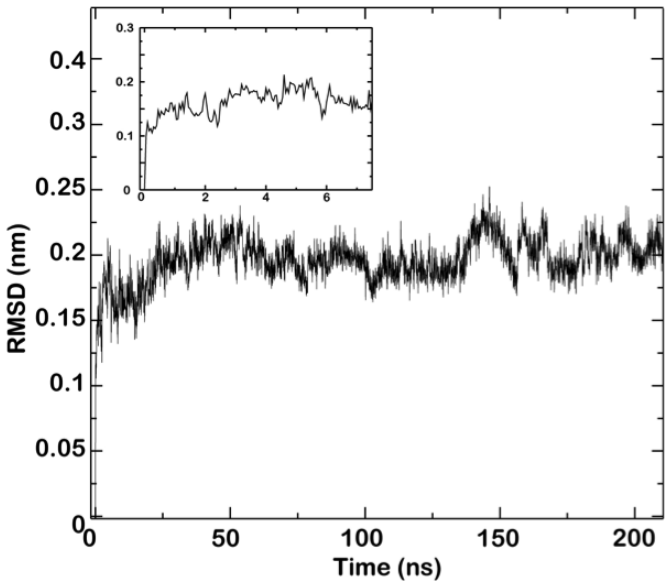


Figure 2.2: RMSD profile of a molecular dynamics trajectory using the initial structure as a reference. The inset shows the changes during the first 5 ns of the trajectory.

where just as before $r_i(t)$ is the position of atom i at time t and denotes the average position of atom i. To do the analysis, the backbone or α carbons atoms are typically selected. The calculation yields large RMSF values for parts of the protein that are highly flexible while portions that are constrained result in lower values. A comparison of these values between a wildtype and mutant simulations can give insight into the effects of mutation or ligand binding. An example of an RMSF plot with the calculations performed on the Cα atoms can be seen in the Figure 2.3. As seen from the corresponding protein structure, large RMSF values result for the loop regions of the protein, but other regions (such as the helix Cα also show large scale motion.

Principal Component Analysis: Principal Component Analysis or PCA is a type of eigenvalue analysis where the complicated dynamics of a system are decomposed into simpler, orthogonal degrees of freedom. PCA is similar to the RMSF calculation discussed above except that the full cross-correlation matrix of all atom pairs is calculated. The eigenmodes of this matrix are determined, and in this way the high frequency, small amplitude fluctuations (small eigenvalues) can be filtered out of the dynamics trajectory, and the larger/slower motions (large eigenvalues) can be extracted. Figure 2.4 shows a projection of a dynamics trajectory on the first two principal modes. The trajectory starts in the right top corner of space spanned by

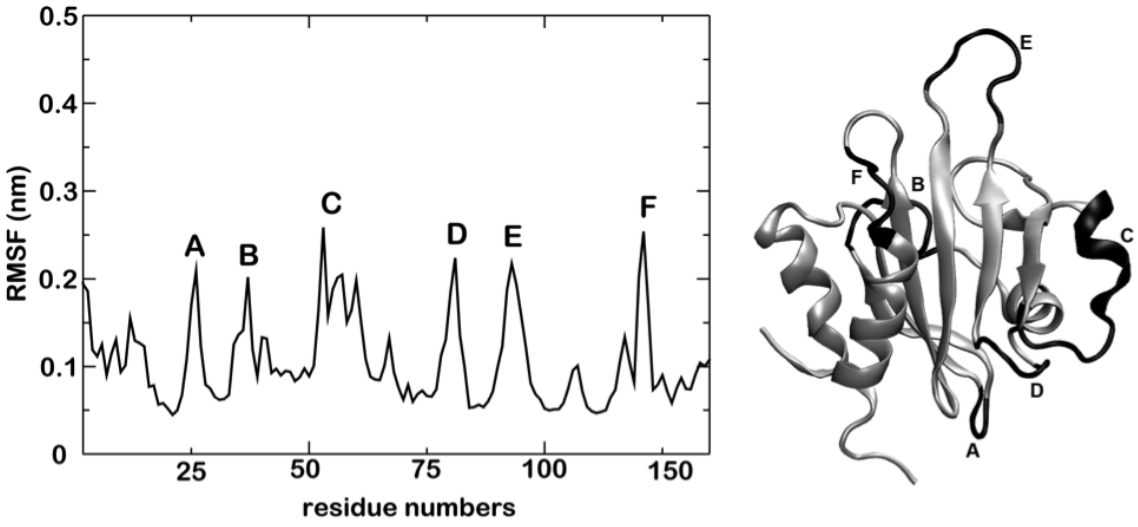


Figure 2.3: RMSF plot for an MD trajectory. The peak RMSF values are labeled and are seen to correspond with the most flexible regions of the protein.

these two modes, migrates to the lower-left quadrant over the first 100 ns, and then remains in this region for the remaining 100 ns. This is the same MD trajectory used in creating the RMSD plot shown in Figure 2.2, but it now suggests that the system does not reach equilibrium for almost 100 ns, not the 50 ns as suggested by RMSD analysis. This underlines the challenges and considerations that one faces is performing these types of simulations and emphasize how carefully the results of an MD simulation should be analyzed.

Molecular dynamics simulations are now being used successfully in a wide variety of chemical, physical and biological systems. As the field progresses, other simulation techniques such as Brownian dynamics, Monte Carlo simulations, and a host of multiscale and coarse-graining methods are emerging. These techniques have some advantages and disadvantages when compared to MD, but one needs to select the correct tool for the problem at hand.

2.4 Thermodynamic Integration Calculations

In molecular dynamics simulation the calculation for the free energy of the transition from the initiate state (e.g. $PROT \cdot Ca$) to the final state (e.g. $PROT \cdot Mg$) is

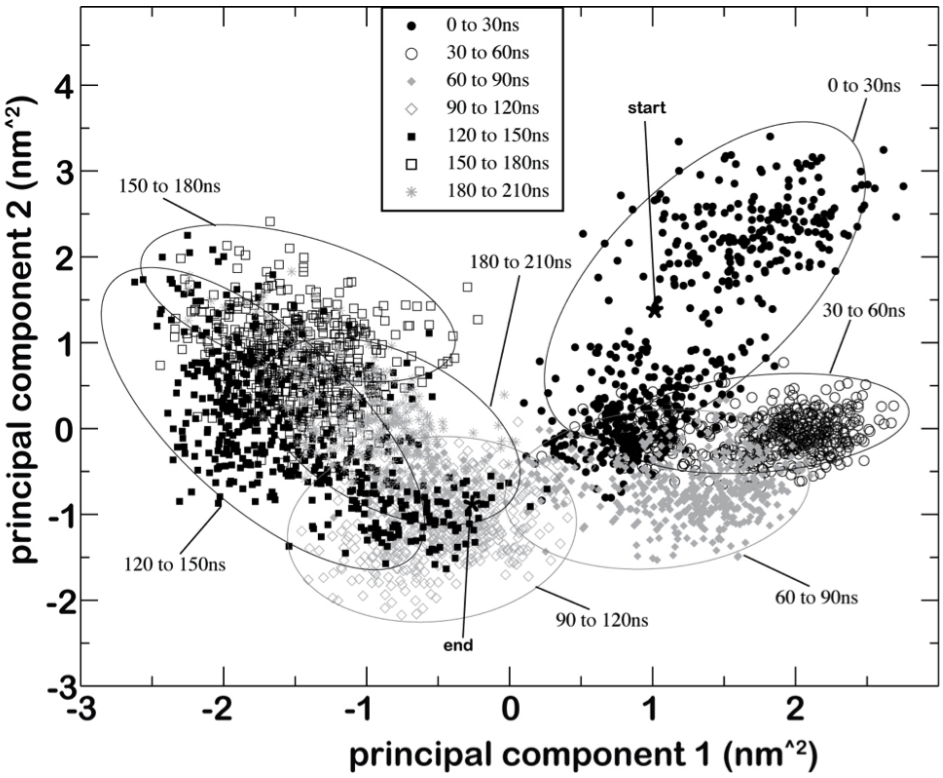


Figure 2.4: Projection of an MD trajectory on the space spanned by first two principal component vectors.

obtained by modifying the energy function, U. A mixed energy function is defined with variable λ.

$$U(\lambda) = (1 - \lambda)U_A + \lambda U_B \tag{2.3}$$

Where U_A and U_B are the energy functions of the initial state A and the final state B. The variable λ is called the coupling parameter which is changed from 0 to 1 in taking from the state A to B. For every λ step the energy derivative is calculated by:

$$\frac{dG}{d\lambda}(\lambda) = \frac{dU}{d\lambda}(\lambda) \tag{2.4}$$

The total change in energy ΔG is obtained by numerical integration of dG/dλ over λ values. The profile is carefully fitted with a polynomial function and then integrated to get the total change energy.

Chapter 3

Effect of D369G mutation on the AC region

3.1 Introduction

D369G, a single residue mutation in the BK_{Ca} channel gene (D369G, mouse; D434G, humans) is known to be linked with epilepsy and paroxysmal dyskinesia in humans (Du et al., 2005). Functionally, the mutation increases the activity of the channel and also the dependence of the channel on the calcium ions. Structurally, the site of mutation lies in close proximity to a key calcium binding residue, D367 and the AC region, a critical component in controlling the calcium-sensitive phenotype of the BK_{Ca} channels (Krishnamoorthy et al., 2005). This indicates that at the molecular level, the mutation might be targeting the calcium-activating machinery of the channel which eventually leads to an increase in the calcium-sensitivity of the channel. The focus of this chapter is to identify the molecular mechanism which is affected by the D369G mutation. To answer that question structural and dynamical studies were performed on AC region, the protein segment which contains the mutation site and also the important calcium activation machinery. The results showed that the mutation brought a reduction in the overall dynamics of the AC region, specifically targeting structural elements which participate in the calcium activation mechanism (Yang et al., 2010).

3.2 Methods

The study was performed by running molecular dynamics (MD) simulations on the AC region protein. Molecular dynamics is a powerful tool for capturing the structural effects produced by a mutation. The structural changes introduced by the mutation are reflected in the conformational states attained by the mutant protein at equilibrium and these states are captured by the simulations. In this study, the MD simulations were performed on the wildtype and the D369G mutant AC region proteins along with a couple of control mutations at location 369. The conformational states of all proteins were characterized and compared to identify the unique effect caused by the D369G mutation. The starting structure for the simulation was obtained by developing a homology model as there was no experimentally obtained structural data available for this channel. Two methodologies were used for running the simulations: The first involved running a single trajectory simulation for 80ns and the second comprised of multiple simulations with the same starting structure but different velocities. The first method which monitored single or transient events of a particular motion, identified the changes caused by the mutation and the second method calculated the statistical significance of the observed change as the protein was able to sample a large number of conformations in this case. Thereafter, correspondence with the functional results was obtained by combining the results with the electrophysiological findings from the Cui lab. The protocol of the simulations is listed below:

1. Homology modeling of the AC region

- Sequence alignment of BK_{Ca} channel protein with the template protein, *MthK* channel
- Prediction of the 3D structure of the wildtype AC region
- Prediction of D369G AC region (epilepsy mutant) and control proteins, D369A, E, N, P and W.

2. Single trajectory molecular dynamics simulations all models

- 80 ns molecular dynamics simulations for each protein with identical simulation conditions.

- Analysis of all trajectories to identify the effect of the mutation
  - Root mean square fluctuation (RMSF) analysis
  - Principal component analysis

3. Statistical significance of the D369G affect

- 10 independent molecular dynamics simulations of each protein with same starting structure but randomly selected initial velocities (A total 0.5 μs of simulations performed)
- Analysis of all trajectories to find the statistical significance of the structural effects
  - Bootstrap root mean squared fluctuation
  - Principal component analysis
  - Hydrogen bond correlation analysis

3.3 Results

3.3.1 Homology Modeling of the AC Region

The homology model for of the AC region was based on the X-ray crystallography data of the *MthK* channel which is a Ca^{2+}-activated prokaryotic channel, with a similar cytosolic gating ring structure as the BK_{Ca} channel (Jiang et al., 2002a). *MthK* channel was chosen as the template because this channel bore functional similarity with the BK_{Ca} channel and also because the sequence of this channel was more than 40% homologous to the BK_{Ca} channel. The homology model was developed using the software MODELLER (Sali and Blundell, 1993) (see the Chapter 2 on Methods for details) and it can be seen in Figure 3.2. Similar models in the previous studies have given important insights into the functional effects of the channel (Shi et al., 2002; Krishnamoorthy et al., 2005; Fodor and Aldrich, 2006).

The sequence alignment (Figure 3.1) showed a gap in the *MthK* sequence at positions 367-370 of the BK_{Ca} channel sequence. This resulted in an insertion of DRDD in

the template structure between the two secondary structure elements, αA and αB (naming scheme from (Jiang et al., 2002b)). The modeling algorithm predicted this insertion as a loop which might be accessible to the solvent. It should be noted that the solvent accessibility of the DRDD loop is important for the channel function if the first residue in the loop, D367 interacts with the intracellular calcium ions. Previous experiments have already shown that D367 might be the calcium binding partner and so the homology model predicted here matches well with those results.

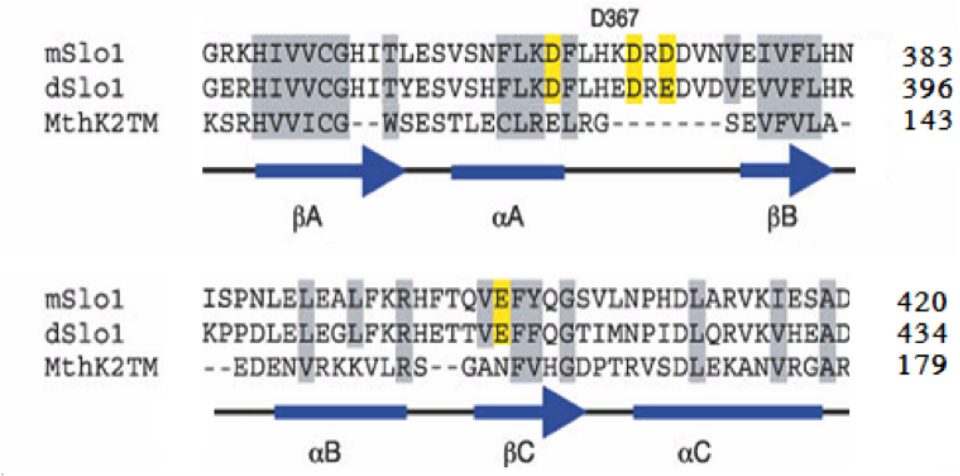


Figure 3.1: Sequence alignment of BK_{Ca} channel AC region (mouse and drosophila) with *MthK* channel. The naming scheme is taken from the *MthK* channel crystal structure. Grey bars represent regions of high similarity.

Next, in order to obtain models for the mutant protein, the side chains at position 369 were changed to Gly, Ala, Glu, Asn, Pro, and Trp by side chain predicting algorithms (Jacobson et al., 2002b). Mutants other than Gly were used as controls because functionally they were not seen to change anything in the channel. The structure of the models obtained for the mutants D369G, D369A, D369E, D369N, D369P and D369W looked very similar to the wildtype structure.

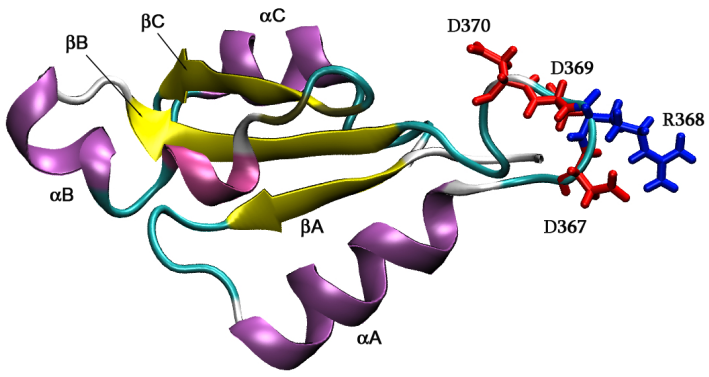


Figure 3.2: Structure of the AC region homology model obtained by using *MthK* channel as a template. AC region stretches from βA to αC in the N–terminus of RCK1 domain. The residue numbers of the DRDD amino acids and the secondary structure elements are labelled. The naming scheme from the *MthK* channel paper has been followed (Jiang et al., 2002b).

3.3.2 Single Trajectory Molecular Dynamics Simulation of all Homology Models

The models obtained above were put through 80 ns of all-atom molecular dynamics simulations. To have an accurate comparison identical simulation conditions were provided to all. The simulations were performed in GROMACS (Hess et al., 2008) using SPC water model as the solvent and Na^+ and Cl^- as counterions. The atomic interactions were evaluated with the OPLS/AA force field (Jorgensen et al., 1996) with a 2 fs time step. Short range non-bonded electrostatics and van der Waals interactions were computed within 9 Å cutoff. Periodic boundary conditions were used and the long range interactions were evaluated using fast Particle–Mesh Ewald electrostatics with grid dimensions of at least 0.12 nm in each direction. Isothermal-adiabatic (NPT) simulations were performed at 1 atm using Parrinello-Rahman extended-ensemble pressure coupling with a time constant of 1 ps, coupled with Berendsen temperature control with a time constant of 0.1 ps. The steps of the MD simulation included an energy minimization until the gradient tolerance was <10 kJ/mol/nm, followed by heating intervals of 50K/20ps until the temperature reached 300 K and then an equilibration run of 60 ps. The production run in each simulation trajectory ran for

80 ns. The snapshots obtained were be analyzed by statistical tools like root mean squared fluctuations, root mean squared deviations, etc to investigate the difference in the structural properties of the proteins.

RMSF Anlaysis

The quantitative comparison of the backbone flexibility among the proteins was obtained by plotting the root mean square fluctuation (RMSF) of the main chain carbon Cα atoms for all four proteins (Figure 3.3). RMSF values calculate the extent of move-

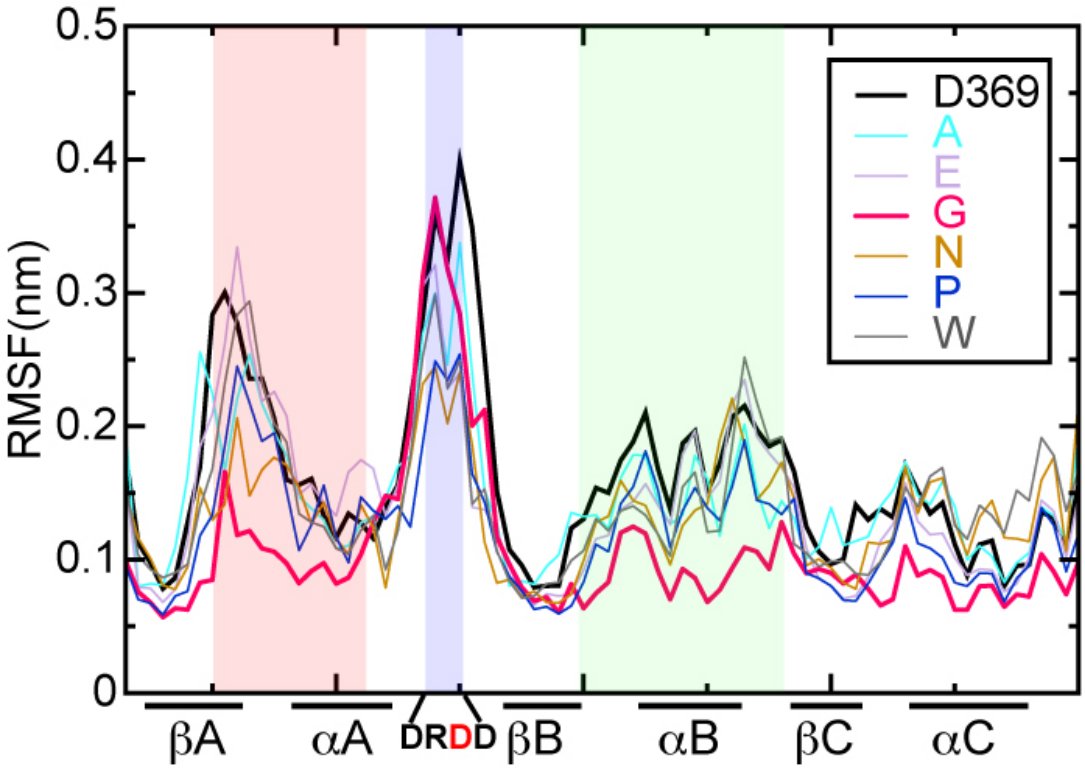


Figure 3.3: RMS fluctuations of the Cα atoms of the WT and D369A, D369E, D369G, D369N, D369P and D369W. Color shades indicate the structural motifs in the AC region, the dynamics of which are significantly affected by D369G.

ment of each residue about its mean position. The parts of the protein that are highly flexible have large RMSF values while the less flexible regions have lower values. We performed the analysis on the Cα atoms of each amino acid. The values obtained are plotted in Figure 3.3. The plots showed that D369G uniquely decreased the RMSF

profile of the AC region while all control mutants mostly followed the profile of the wildtype AC region. Interestingly, the D369G mutations caused a global change in the dynamics of the entire AC region rather than having a local effect in the DRDD loop, the site of mutation. The largest difference was seen in the αA and αB helices indicating an allosteric connection between the DRDD loop and the αA helix and the αB helix.

Principal Component Analysis

To further investigate the dynamics of the AC region, the full dynamics trajectory was decomposed into simpler orthogonal modes using principal component analysis (See Chapter 2 for details on PCA). Figure 3.4 shows the dynamic range of the largest

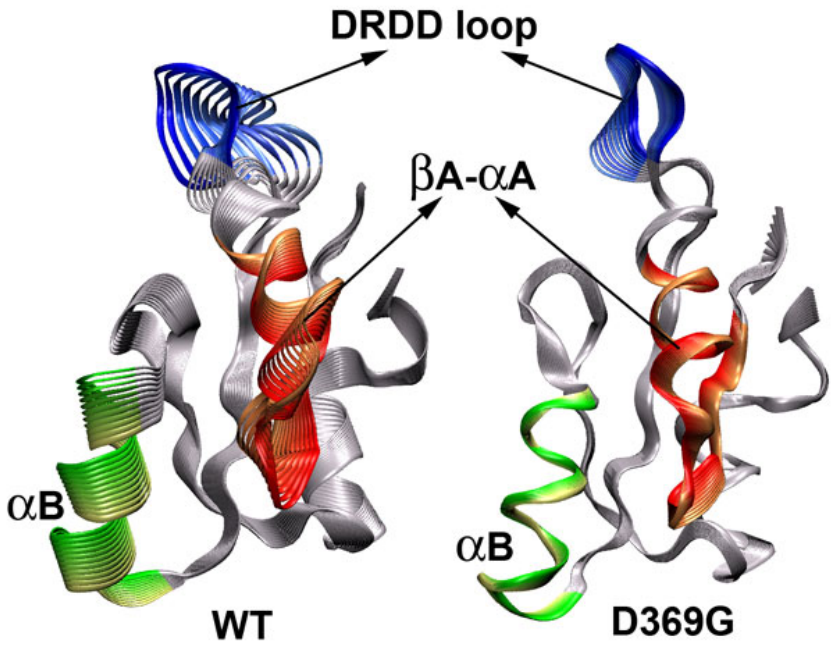


Figure 3.4: Range of motion for wildtype and D369G protein along their principal modes. Secondary structural elements include αA (red), αB helix (green) and the DRDD loop (blue).

amplitude mode of each protein. The amplitude of the most significant motion in the

wildtype protein was seen to be much larger than that of the D369G mutant. In the principal mode of the wildtype protein, the dynamics predominantly comprised of the local movements of the αA (red), αB helix (green) and the DRDD loop (blue). The dynamics in the principal mode of the mutant protein involved just the DRDD loop motion. This revealed that the dynamic nature of the wildtype protein was lost upon making the mutation and the entire protein started to demonstrate a restricted motion.

3.3.3 Statistical Significance of the D369G Effect

This purpose of this section was to find the statistical significance of the change in dynamics observed after making the D369G mutation in the AC region. The previous section presented results from a single trajectory simulation of 80 ns, which might have insufficiently sampled the equilibrium states. Protein simulations deal with systems of multiminimum character and there are energetic barriers which may not be crossed sufficiently frequently in accessible computation times. Also, different properties of the system take different times to achieve convergence. A single trajectory starts with one set of initial conditions and this sometimes can cause the initial biases to be carried over for a long time, causing the system to get stuck in a local minimum and preventing the properties to converge. A number of techniques have been developed for improving the sampling of molecular dynamics simulations and they have been successfully applied to several proteins. One of these techniques, which involved running multiple independent short simulations with different initial velocities was applied in this section. Several independent simulations were run for both the wildtype and the mutant AC region and the snapshots obtained from all simulations were characterized for the mutation effect.

Simulation Protocol

A total of 10 independent MD simulations of 50 ns duration were started from the atomic level structures of both AC regions, obtained earlier by homology modeling and single residue mutation on D369 to G. Each simulation was provided a unique

initial velocity picked from the temperature dependent Maxwell Boltzmann distribution. This method produced more number of conformations in the same time as the simulations were run in parallel. Every simulation consisted of 80 residue AC region, Na^+ counterions and 10 Å water shell surrounding the protein on each side, with a total of ∼20,000 atoms. Using parallelized MD program GROMACS, all simulations were performed with OPLS/AA force field and SPC water model. Interactions were evaluated using a time step of 2 fs using a leap-frog md integrator. Short range non-bonded electrostatics and van der Waals interactions were computed within 9 Å cutoff. Periodic boundary conditions were used and the long range interactions were evaluated using fast Particle-Mesh Ewald electrostatics with grid dimensions of at least 0.12 nm in each direction. Isothermal-adiabatic (NPT) simulations were performed at 1 atm using Parrinello-Rahman extended-ensemble pressure coupling with a time constant of 1 ps, coupled with Berendsen temperature control with a time constant of 0.1 ps. The steps of the MD simulation included an energy minimization until the gradient tolerance was <10 kJ/mol.nm, followed by heating intervals of 50K/20ps until the temperature reached 300 K and then an equilibration run of 60

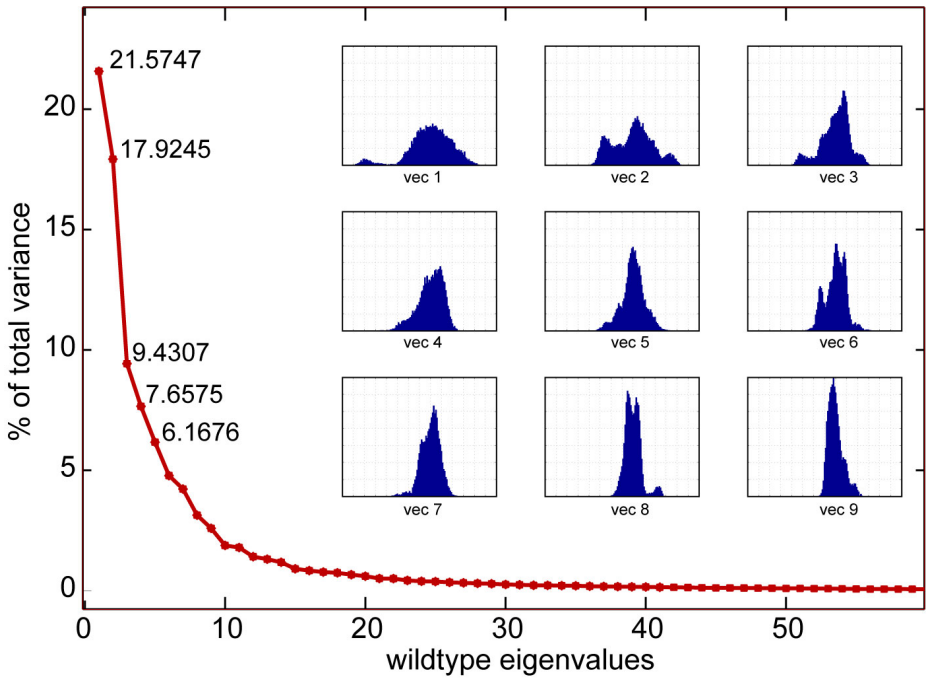


Figure 3.5: Eigenvalue variance of the entire wildtype basis set obtained the PCA all wildtype trajectories (concatanated). The inset shows the projections of the full trajectory on eigenvectors 1 to 9.

ps. The production in each simulation trajectory ran for 50 ns amounting to 0.5 μs of the combined trajectory.

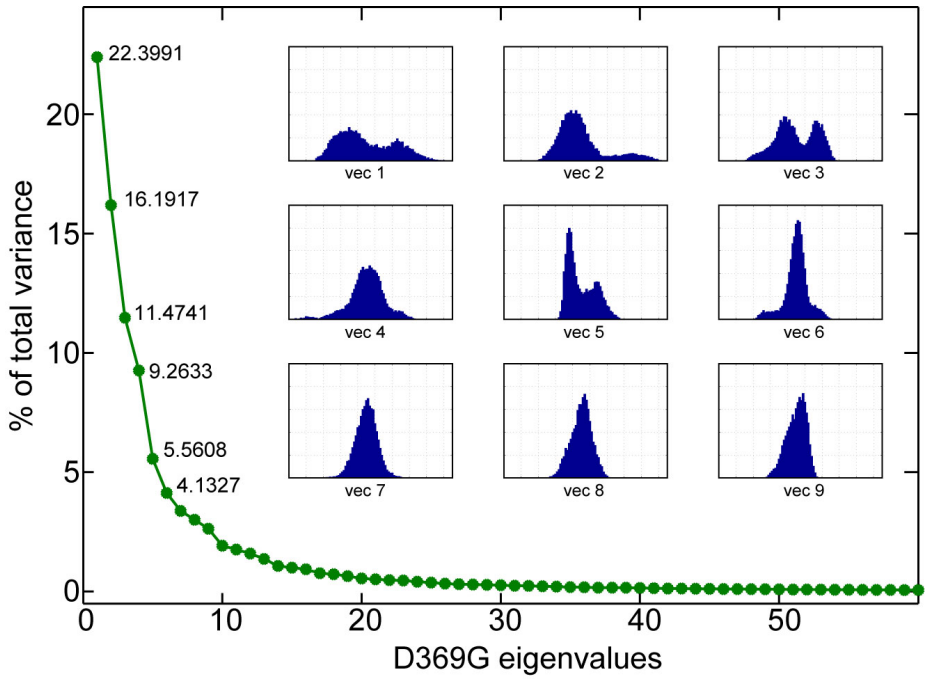


Figure 3.6: Eigenvalue variance of the entire D369G basis set obtained the PCA all D369G trajectories (concatanated). The inset shows the projections of the full trajectory on eigenvectors 1 to 9.

Sampling of the Conformational Space by Multiple Simulations

To gain insight into the structural changes occurring in the protein, the conformational space of the molecule was examined by means of principal component analysis (PCA, see the Chapter 2). The Cartesian coordinates of the backbone atoms was used to obtain a covariance matrix of atom fluctuations about their mean positions. This defined the conformational space of the protein backbone with 3N degrees of freedom (N=number of atoms). The conformational space of the ensemble of trajectories was combined into one, and the resulting matrix was diagonalized to generate the orthogonal modes or the eigenvectors and the eigenvalues of the data. These vectors were ranked on the magnitude of the eigenvalues, with the first mode accounting for the most significant motion of the protein. The projection of the entire trajectory on each mode showed that the initial modes displayed a somewhat bimodal distribution

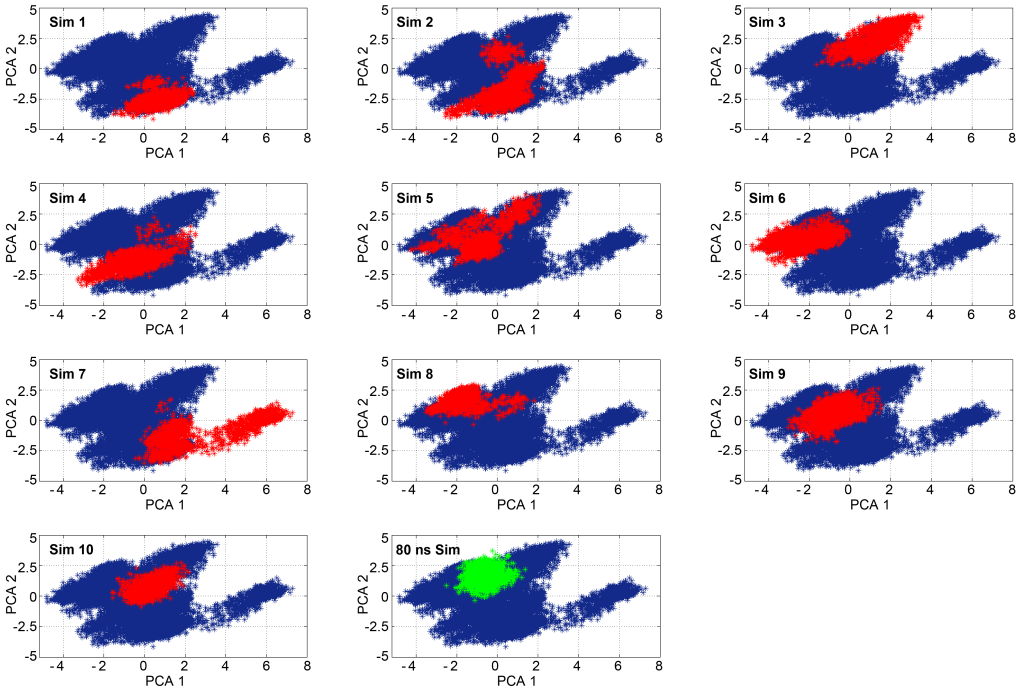


Figure 3.7: Projections of the individual wildtype trajectories on the first two eigenvectors of the entire wildtype basis set. For each graph the blue dots represent the concatenated set and the red dots are the individual trajectories. The green projection shows the 80 ns trajectory from Section 3.3.2 on the entire basis set.

which became progressively sharper for the higher modes, demonstrating their decreasing variance (Figure 3.5). Examination of the protein motion corresponding to each mode showed that starting modes (with higher eigenvalues), usually represented the long range motions in the protein and the higher modes (with smaller eigenvalues) captured the short range motions or the bond vibrations. For the AC region system, the first three modes in both the wildtype and the mutant protein, accounted for ∼50% of the variance of fluctuation (21.57%, 17.92%, 9.43% - wildtype; 22.39%, 16.19%, 11.47%) (Figure 3.5 and 3.7).

The progression of each individual trajectory on the conformational space was monitored by examining their projection on the two most significant modes (Figure 3.7). Each trajectory was observed to sample different parts of the conformational space. Though, each started with the same structure, they all ended up in a different region. Some of these regions overlapped with each other, but most of them were unique. The single 80 ns run from the previous section was found to be confined to a small

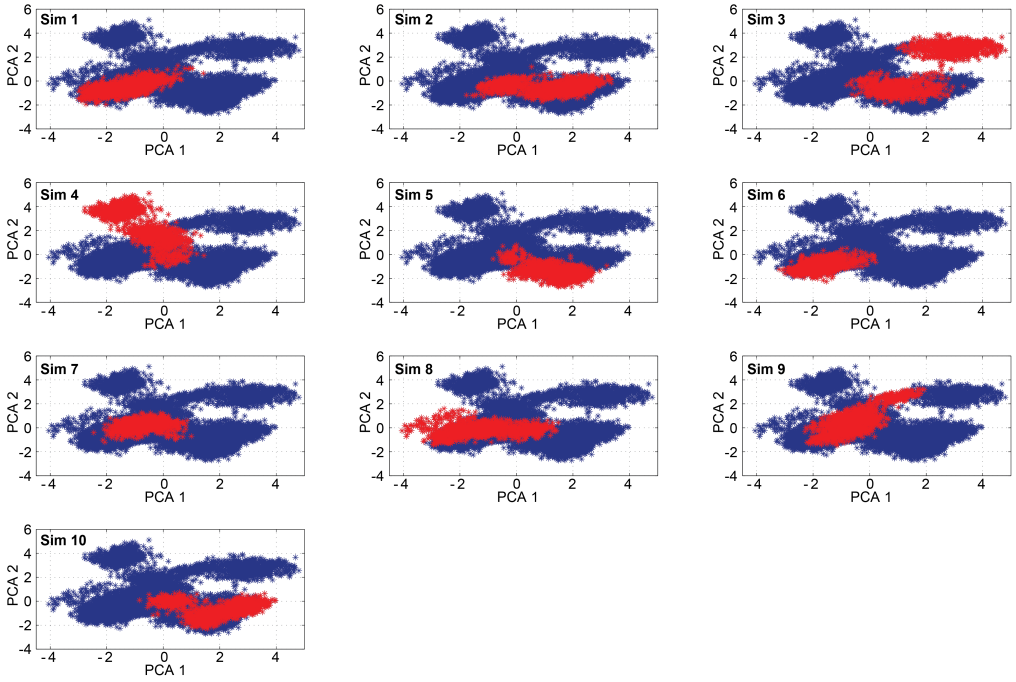


Figure 3.8: Projections of the individual D369G trajectories on the first two eigenvectors of the entire D369G basis set. For each graph the blue dots represent the concatenated set and the red dots are the individual trajectories. The green projection shows the 80 ns trajectory from Section 3.3.2 on the entire basis set.

space indicating that the multiple trajectories were able to capture many more states which might have been inaccessible in a single run. This phenomenon was seen in both the proteins confirming that the method was able to provide a better statistics.

Comparison between Structural Changes in Wildtype and Mutant Proteins

The structural changes introduced by the D369G mutation were identified by comparing the projections of the wildtype and the mutant trajectories on the same basis set (Figure 3.9). The wildtype basis set was used as the reference. The comparison showed that the mutant protein spanned a smaller area on the conformational space as compared to the wildtype trajectory. The density of occurrence of the different conformational states was obtained by binning the projection data over time. The peaks corresponded to these states and the height gave the frequency of their

occurrence in each trajectory. The results showed that the wildtype plot consisted of many peaks with smaller rate of occurrence for each state. The mutant trajectory, however, had just two broad peaks with a very high frequency. This indicated that the wildtype protein sampled through many significant conformational states, revealing the dynamic nature of its backbone. The mutant protein, on the other hand, preferred fewer conformational states and kept sampling close to the preferred conformations. Thus, from these results it became clear that the mutation indeed changed the dynamical nature of the protein and made it rigid.

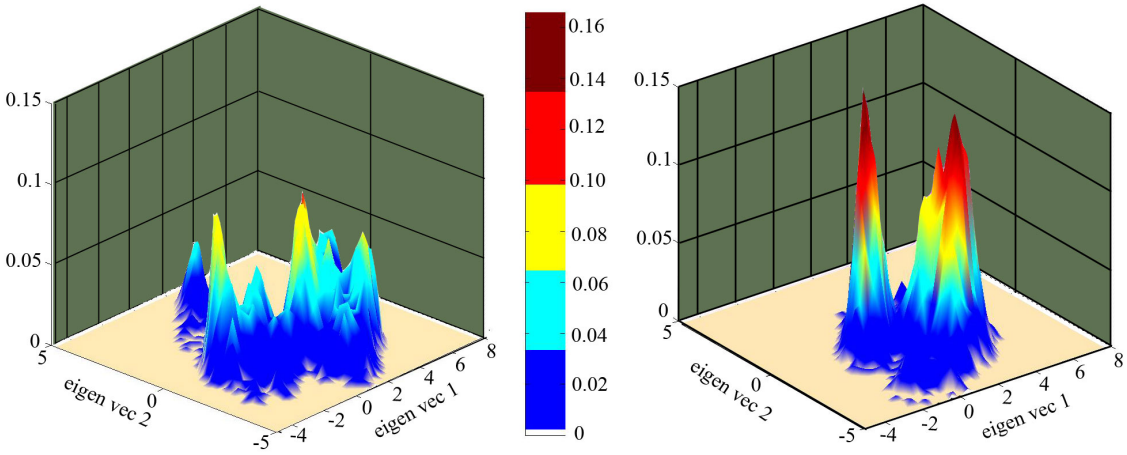


Figure 3.9: Projections of wildtype (left) and D369G (right) trajectory on the 1st and 2nd eigenvectors of the wildtype basis set. The x and y–axis represent the projections on the vectors 1 and 2 and the vertical axis represents the fraction of the trajectory lying in the projection value bin. The higher number of peaks in wildtype plot shows that it explores more number of conformational states.

Bootstrap Root Mean Squared Fluctuations

To quantify the flexibility of the protein and to identify the structural elements specifically affected by the mutation, RMSF calculations were done on the trajectories. Customarily, the RMSF calculation is done on a single trajectory but in the present case the data from the ensemble of trajectories were combined by a bootstrap analysis on their RMSF values. This analysis generated a sampling distribution of the RMSF values for each residue. First the number of independent points in the MD trajectories were found by looking at the autocorrelation of the root mean square deviations of the structures. This analysis gave a correlation time of 1ns, indicating that there

were 50 independent structures in each of the 50 ns trajectories. To calculate the RMSF, 50*10 random points were selected with replacement, repeating 100 times, yielding a mean RMSF and a standard error of mean for each amino acid. The significant difference of values between the mutant and the wildtype proteins were found by applying Student's t–test. Contiguous regions of four or more amino acids that showed significant difference at a significance level of $p < 0.005$ were identified as the affected regions. The standard error of mean obtained from the bootstrap analysis is plotted as the width of the RMSF value in the Figure 3.10. The regions identified by the Student's t–test are shown by shading.

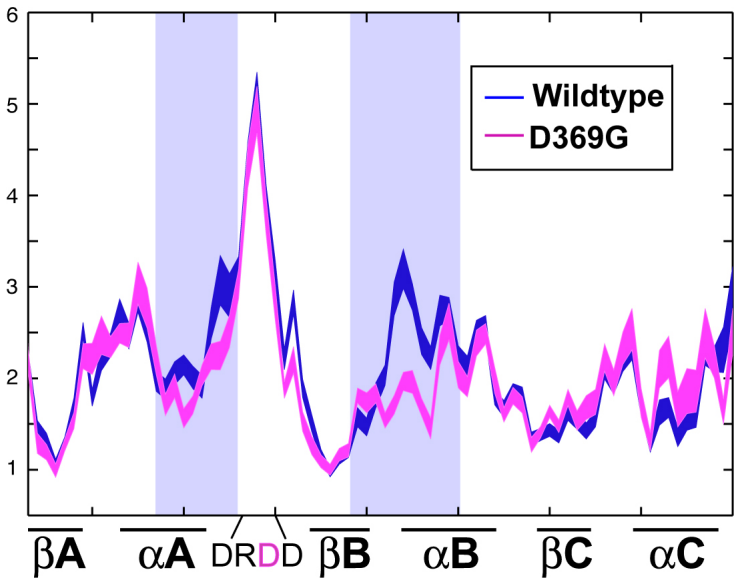


Figure 3.10: RMSF values obtained after bootstrapping data from the multiple trajectories. The width of each plot represents the standard error of mean of the RMSF value for each residue. The regions significantly different in the D369G protein from the wildtype protein are shaded light blue.

The results showed that for most of the residues the mutation did not have any effect on the dynamics. This included the site of mutation, the DRDD loop lying between helices αA and αB. The regions which showed significant difference in dynamics were helices αA and αB, highlighted by blue bars. These results were consistent with the outcome of the single trajectory molecular dynamic simulations seen in Section 3.3.2. Thus, this confirmed the structural consequence of the D369G mutation as being a loss in dynamics of two helices in the AC region protein. Interestingly, these helices

are located several residues downstream of the mutation site which opens a possibility of an allosteric connection between the loop and the helix.

Hydrogen Bond Interactions

A method introduced by Bradley et al. (2008) was used to monitor the making and breaking of the non-covalent bonds during the simulation. The method involved making a list of all possible non-covalent bonds in the protein and then calculating the frequency of their formation. The statistics of these contacts over the entire trajectory provided a list of contacts which played an important role in the protein dynamics. This list mostly yielded the important hbond contacts such as helix i, i+4 interactions which were formed 100% of the time. In order to identify the dynamical contacts and the correlation between them, a correlation coefficient C_{ij} was calculated for every i,j pair of contacts (Details can be found in the Methods 2). Then by using chi-squared statistics the contacts were classified as significant at the 95% confidence level. The method yielded a long list of contact pairs which were then clustered using UPGMA clustering to produce groups of contacts which were significantly correlated to each other. Each group when viewed on the structure displayed a network of contacts connecting different regions of the protein. To find how the interactions between the functionally important regions were affected by the mutation, the list was screened for groups which connected the calcium binding residue D367 to other regions of the protein. These contacts were then mapped on the structure and can be seen in Figure 3.11. The wildtype protein displayed strong correlations between the contacts in the DRDD loop and in the αA and αB helices. The mutant protein showed different correlations involving hydrogen-bonds within the helices which were mostly required to improve the rigidity of the helices. The principal component analysis results showed that the wildtype protein was more dynamic and the result here revealed that the contact fluctuations in different regions of the wildtype protein were also correlated to each other confirming the correlated motions in the protein. The mutant protein displayed a lack of these correlations and therefore does not demonstrate large motions.

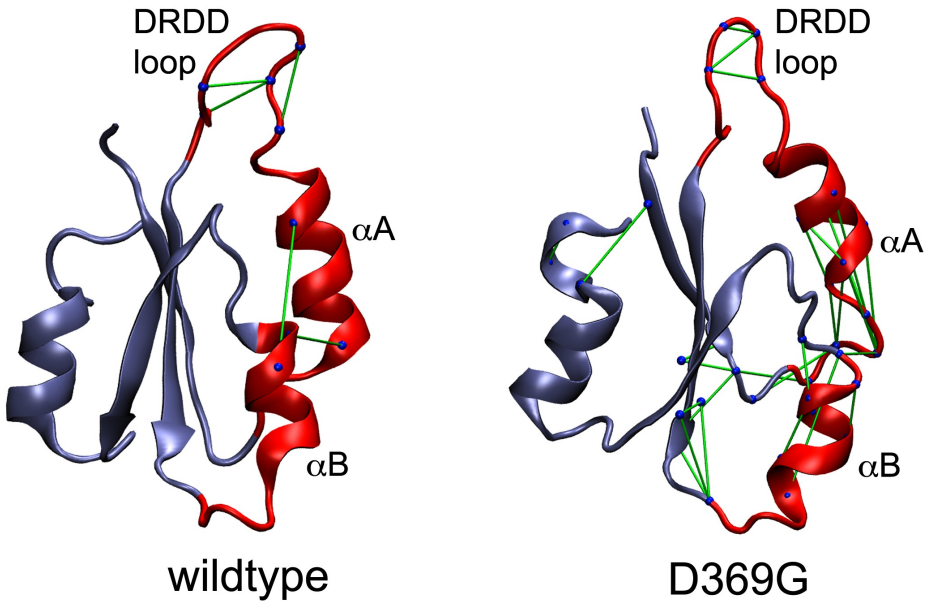


Figure 3.11: The average structure of each protein is shown here with green lines representing the significant correlations between the secondary structure elements. The significance of the correlations were obtained from the non-covalent analysis as discussed in section 3.3.3.

3.3.4 Conclusions

MD simulations on the AC region revealed the structural consequence of the D369G mutation effect. The D369G mutation was found to reduce the overall dynamics of the AC region with the loss in flexibility localized in the secondary structural elements αA and the αB. The local movements of these motifs appeared to be correlated with the movement of the DRDD loop in the wildtype simulation but they were lost in the mutant case. The dynamics of the all other mutants, which were used as controls, followed the wildtype profile. This change in dynamics was further confirmed by sufficiently sampling the protein conformational space and the results confirmed that the mutation indeed caused a reduction in the overall dynamics of the AC region.

3.4 Relationship with Channel Function

3.4.1 Relation with the gating mechanism of the channel

To understand the implication of the change in dynamics in the AC region on the gating mechanism of the channel, homology model of the membrane-spanning pore was built and superimposed with the AC region on a full channel template. *MthK* channel structure was used for the template. The resulting model (Figure 3.12) revealed that the AC region was located beneath the pore with the DRDD loop lying at the periphery. The αA and the αB helix, which experienced the maximum change in their backbone flexibility by the mutation, were situated directly beneath the pore suggesting a possibility of interaction with the pore. This strongly suggests that any change produced in the dynamics of these regions could directly affect the dynamics of the pore.

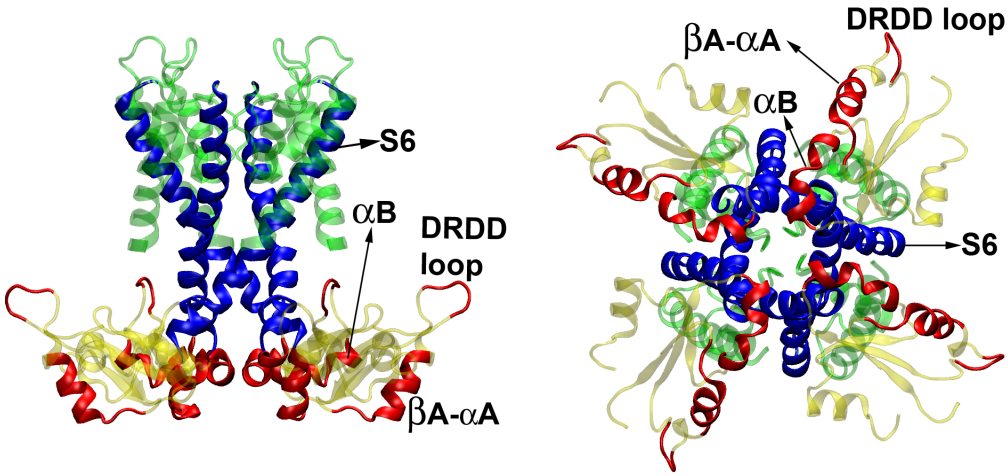


Figure 3.12: Homology model of the pore based on the Kv1.2 channel and the AC region model based on the *MthK* channel were superimposed on the *MthK* full channel template. The superimposition of both models shows the location of the AC region with respect to the pore.

3.4.2 Correspondence of the Simulation Results with Experiments

Combination of the structural results from the simulations and the functional results from electrophysiological findings (data obtained from Cui lab) suggested that the decrease in flexibility caused by the insertion of the Gly residue in the DRDD loop went hand in hand with the change in Ca^{2+} related function of the channel. Ala mutations were done by the Cui lab on each residue of the AC region to find its effect on the Ca^{2+} related function of the channel. This experiment picked out 10 residues (labeled by red, blue and green colors in Figure 3.13) which significantly altered the function of the channel indicating that they participated in the mechanism in some way. These residues were present in the DRDD loop, αA, αB and βC motif. The simulations revealed an allosteric connection between D369 in the DRDD loop and the distal regions αA and αB. Ala scanning suggested that the aforementioned structural regions participated in the calcium activation mechanism of the protein. Combination of the two results revealed that the epilepsy mutation was reducing the dynamics of those motifs which contained residues participating in the calcium activation mechanism of the channel.

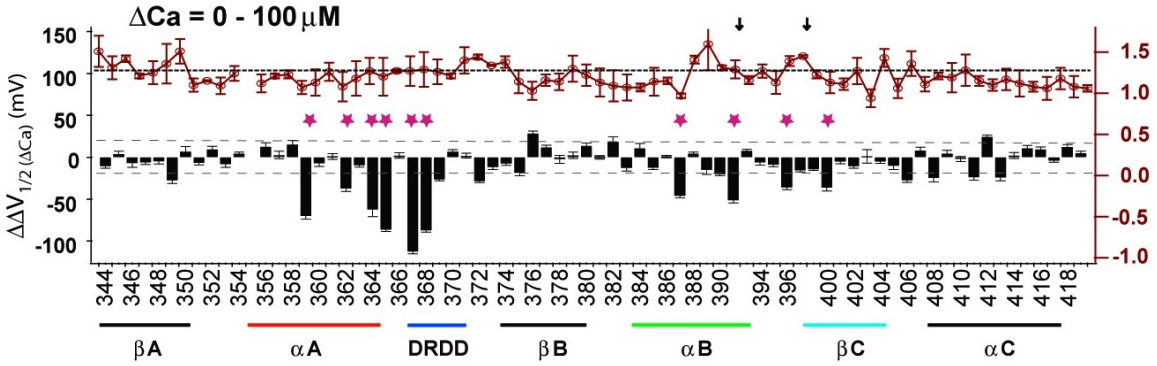


Figure 3.13: Ala scanning of the AC region residues to find the change in calcium sensitivity. The residues were individually mutated to Ala and that reduction of calcium sensitivity from the wildtype channel was recorded as $\Delta\Delta V_{1/2}(\Delta Ca^{2+}) = \Delta V_{1/2}^{WT} - \Delta V_{1/2}^{Mutant}$, where $\Delta V_{1/2} = V_{1/2}$ at 99.3 to 111.5 mM $[Ca^{2+}]_i - V_{1/2}$ at 0 $[Ca^{2+i}]_i$. Mutations with $\Delta\Delta V_{1/2}(\Delta Ca^{2+})$ falling beyond the $\pm$ 20 mV (dashed lines) interval significantly affect Ca^{2+i} activation, which are (denoted by asterisks) F359A, D362A, L364A, H365A, D367A, R368A, L387A, F391A, T396A, and F400A. Structural motifs are indicated by horizontal thick lines. Data obtained from the Cui lab.

3.4.3 Correspondence of the Model with the Crystal Structure

All simulations shown above were based on the homology model of the AC region as the crystal structure of the protein was not available at the onset of this research. Recently, two groups (Yuan et al., 2000; Wu et al., 2010) released crystal structures of the BK_{Ca} channel which provided the opportunity to validate the homology model. The model was superimposed on the AC region of the Ca^{2+} free crystal structure and it gave an RMSD of 1 Å. Figure 3.14 shows the homology model overlaid on the crystal structure.

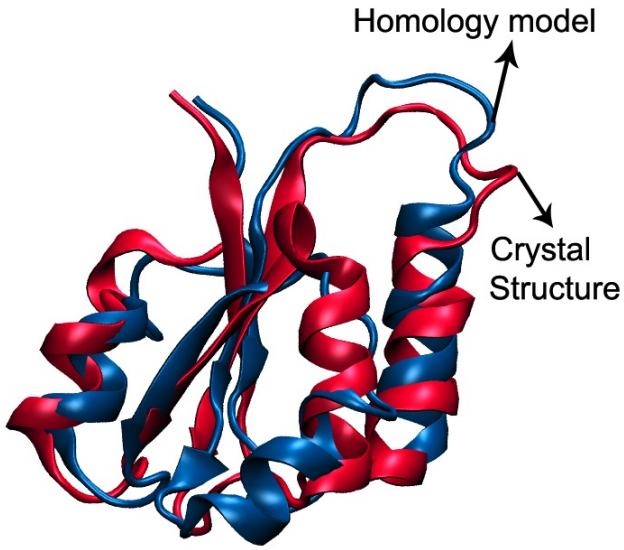


Figure 3.14: Superimposition of AC region homology model (blue) on the (Du et al., 2005) crystal structure (red). The homology model bears striking resemblance to the crystal structure confirming the accuracy of the model.

3.5 Discussion and Conclusions

The D369G mutations in the BK_{Ca} channel had previously shown a marked increase in the open-channel probability of the channel and a $3-5$ fold increase in the Ca^{2+} sensitivity (Du et al., 2005). The results shown here explain the structural consequence of this mutation. AC region, a small segment in the N–terminus of the RCK1

domain, was found to be directly influenced by the mutation. The protein segment experienced a loss in backbone flexibility with the effect localized in structural motifs, αA and αB.

Given the challenges of the structural studies, molecular dynamics simulations provided the best means to fully understand the molecular mechanism. With the use of the simulations The simulations allowed full characterization of the conformational states attained by each protein and gave an explanation of the structural effect produced by changing a bulky Asp to a small Gly. Each protein presented a different behavior. The wildtype protein was very dynamic and it hopped around in many different states. The mutant protein, preferred to keep sampling similar states throughout the length of the simulation. This was seen in both sets of simulations, the single trajectory run of 80 ns and the multiple trajectory runs of 0.5 μs confirming that the observations were the inherent characteristics of the two proteins. Predictably, mutation from Asp to Gly caused a change in the backbone flexibility, but something not expected was that the change was spread through the entire protein. The mutation was made in the DRDD loop but the effect was seen couple of residues downstream in motifs αA and αB, indicating the allosteric nature of the effect.

The single simulation for 80 ns was able to discern the effect of the D369G mutation but in order to be statistically confident of the observed changes it was necessary that we allowed the proteins to sample a larger conformational space. We used an approach where instead of running a single molecular dynamics calculation; we ran an ensemble of simulations, each starting with different velocities picked from random distributions. We found that by using this method we were able to access a variety of conformations including the ones that had relatively shorter lifetimes and did not appear in the single run calculation. Although, in principle we could have sampled those rare conformations in the single run, but such a simulation would have to be extremely long and therefore it would not have been very efficient. By using this method we are able to explore a larger conformational area in less computational time. The large dataset obtained with this new method, was able to better represent the full range of motion experienced by the protein. Further, by using bootstrapping algorithm on the ensemble of trajectories we are able to obtain statistically significant values. The bootstrapping not only measured the accuracy of the data, but also

yielded the sample error of the mean, which is the closest approximation of the population mean.

Combination of the structural results from molecular dynamics with the functional results from the electrophysiology experiments found the structure-function relationship of this mechanism. Structural studies were able to identify the motifs affected by the D369G mutation and the mutagenesis experiments recognized that these motifs had key roles in the calcium related function of the full channel. To further confirm this connection, Cui lab performed more experiments combining the D369G mutation with the mutations in these key motifs (Yang et al., 2010) and found that mutations in these motifs indeed altered the functional effect produced by the D369G mutation. This is a noteworthy example of how the integration of two types of studies was able to throw a great deal of light on the understanding the structure-function relationship of the effect of a mutation.

It is encouraging that the results obtained from the homology model of the AC region correlated very well with the functional effects. On comparing our model with the crystal structure released this year we found that our model was indeed very similar to the crystal structure. Thus, it emphasize the fact that the computational protein prediction algorithm in use today are very powerful and can be used confidently for gaining insights into the structure and function of those proteins whose structural data is challenging to obtain experimentally.

Chapter 4

Effect of D369G on the RCK1 domain

4.1 Introduction

Recently, two crystal structures of the BK_{Ca} channel gating ring have been experimentally solved revealing two states of the channel protein (Yuan et al., 2010; Wu et al., 2010). 3MT5 crystallized in the presence of 50 mM calcium concentration shows a calcium ion coordinating with the calcium bowl binding site. The other structure, 3NAF is ion free and is considered as the the closed state of the channel. None of these structures show calcium binding in the RCK1 domain, but based on the closed state of the channels, the authors predict that there might be a high affinity calcium binding site close to residue D367. From the structures, it appears that the residues from the loop between αA and βB, the loop between αG and αH and the βG strand might enclose a binding site. Mutagenesis studies have given strong evidence that residue D367 at this location might be a calcium binding residue (Xia et al., 2002) and recently (Zhang et al., 2010) found a resideu in the βG motif, E535 which could act as another coordinating partner. They also found an indirect role of R514 (downstream of αG) in this mechanism.

The previous chapter talked about the effect of the D369G mutation on AC region, a critically important protein segment. AC region is the 80 residue segment in the N-terminus of the RCK1 domain stretching from βA to αC. In the light of the recent results, it looks like the structural motifs αG, αH and βG might play a major role in the calcium activation mechanism and thus would also experience the effect of the

mutation. AC region might reveal part of the story and so in order to get a deeper understanding into the effect of the D369G mutation, it is important to include the aforementioned regions in the study.

This chapter presents the results on the structural and dynamical effect of the D369G mutation on a single subunit of the BK_{Ca} channel, comprising of RCK1 and RCK2 domains. MD simulations were performed on the wildtype and the D369G mutant proteins and it was observed that the D369G mutation still affected the global dynamics of the protein structure but this time it was seen on a much larger scale. Wildtype protein exhibited that it existed in two major conformations and the mutation was seen to drive the protein to a single conformation, which is more stable and less dynamic.

4.2 Methods

The crystal structure, 3MT5 was used as the starting structure to run the MD simulations on the full subunit. The multiple simulations methodology used in previous chapter was advantageously used here to sample a large conformational space. A series of 12 independent 50 ns runs (total of 0.6 μs) were run for both the wildtype and the mutant protein, starting with different initial velocities. The protocol of the simulations is listed below:

1. **Preparation of the starting structure**
    - Completed the crystal structure sequence by predicting the missing residue locations via secondary structure prediction algorithms
    - Performed mutation at residue D369 to G to obtain the mutant starting structure
2. **Molecular dynamics simulations of the two proteins**
    - 12, 50 ns Molecular dynamics simulations starting with different velocities for each set (total of 0.6 μs)
    - Analysis to identify the structural and dynamical effects of the mutation
        - ◇ Principal component analysis

- ◇ RMSF, RMSD
- ◇ Hydrogen bond and salt bride interactions

4.2.1 Preparation of the Starting Structure

The crystal structure 3NAF (calcium free form of BK_{Ca} channel gating ring) was used for this study (Wu et al., 2010). A single subunit of the gating ring was used as the starting structure. The structure had missing coordinate information for residues 290 to 330, 667 to 683 and 834 to 871 which were predicted by using the ab-initio structure prediction algorithm of ROSETTA (Rohl et al., 2004a; Simons et al., 1999b). To obtain the mutant protein, the residue at position 369 was replaced by glycine using the Protein Loop Structure Predicting algorithm PLOP (Jacobson et al., 2004).

4.2.2 Molecular Dynamics Simulation

For the simulations the starting structure was solvated in TIP3P water with a 10 Å padding on each side. Na^+ CL^- counter–ions were used to maintain the salt concentration at 50 mM. Using parallelized MD program Gromacs 4.1 (Hess et al., 2008), all simulations were performed in an NPT ensemble with OPLS/AA force field and TIP3P water model. Forces were calculated by a leap-frog stochastic dynamics integrator at a time step of 2 fs. Short range non-bonded electrostatics and van der Waals interactions were computed within 9 Å cutoff.. Periodic boundary conditions were used and the long range interactions were evaluated using fast Particle-Mesh Ewald electrostatics with grid dimensions of at least 0.12 nm in each direction. Isothermal-adiabatic (NPT) simulations were performed at 1 atm using Parrinello-Rahman extended-ensemble pressure coupling with a time constant of 1 ps, coupled with Langevin dynamics temperature control. The inverse fraction constant for the temperature control for each group was set at 2 ps, since this results in a friction that is lower than the internal friction of water, while being high enough to remove excess heat. The steps of the MD simulation included an energy minimization until the gradient tolerance was <10 kJ/mol/nm, followed by heating intervals of 50K/20ps until the temperature reached 300 K, and then an equilibration run for 60 ps. The production in each simulation trajectory ran for 50 ns amounting to 0.6 μs of the

combined trajectory. For each protein, two simulations, Sim1 and Sim2 were first started which ran for 50 ns to allow the structures to further equilibrate and get rid of any crystal structure artifacts, if present. Thereafter, the end structures of these two simulations were used to generate 5 different simulations from each, Sim1a-e and Sim2a-e, amounting to 12 simulations for each protein. All sets had identical simulation conditions except for the initial velocities. The velocity for each simulation was randomly picked from the Maxwell-Boltzmann distributions and assigned at the beginning of the simulation.

4.3 Results

4.3.1 RMSD

The simulations were started with the crystal structure which might carry some initial biases. To ensure that the system was equilibrated enough, the evolution of each trajectory was monitored as it went away from the crystal structure. For the two initial trajectories Sim1 and Sim2, an initial rapid rise was observed in the RMSD values which leveled off after around 10 ns. This initial change in the RMSD values contained fluctuations which might be correlated to each other and hence can be attributed to the equilibration phase of the simulation. Thus the initial 10ns of trajectories Sim1 and Sim2 were removed from further analysis. The 10 trajectories, which were generated from the end points of the first two simulations (Sim1a-e and Sim2a-e), showed all RMSD values in the equilibrated region, with continued fluctuations of 1-2 Å indicating that they had reached equilibrium.

4.3.2 Characterization of the Wildtype Protein

To gain insight into the structural and dynamical features of the wildtype protein principal component analysis was performed on all trajectories. The 12 different trajectories were combined to obtain an overall conformational space and a collective covariance matrix of all atomics fluctuations. Diagonalization of this matrix produced eigenvectors and eigenvalues which were ranked based on the magnitude of the

eigenvalues. The rankings showed that the first 9 vectors accounted for around 85% of the entire motion of the trajectory. This allowed the reduction of the degrees of freedom of the system to the most significant motions of the protein. The other 15% of the motions mostly comprised of the bond vibrations and short length fluctuations which might not be important for the function of the protein.

Next, the entire trajectory was projected on each of the selected vectors to observe the behavior of the protein along those vectors. The distribution of the projections on each mode gave an idea of the extent of fluctuations occurring along that mode. The first mode showed the largest variation and the spread became smaller and smaller as the rank increased, indicating their decreasing contribution to the total variance (Figure 4.1).

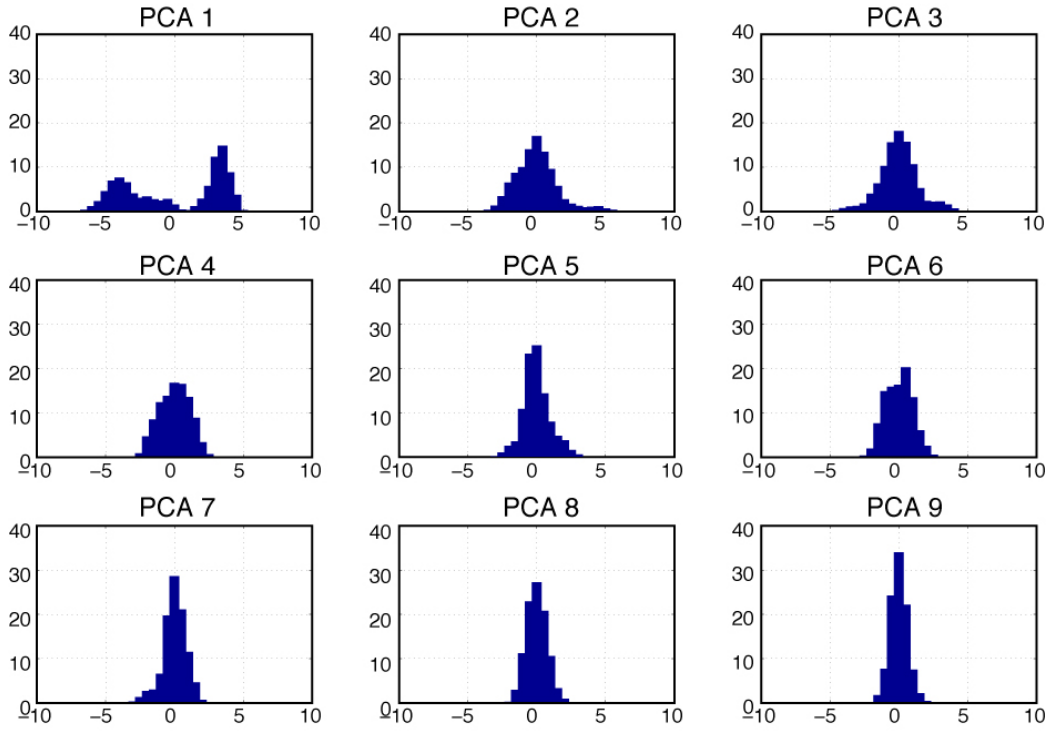


Figure 4.1: Distribution of the projections of the wildtype trajectory on PCA vectors 1 to 9. The first mode represent the largest variation. The higher modes become smaller indicating their decreasing contribution to the total variance.

Sampling of the Multiple Wildtype Trajectories

Projection of each trajectory on pairs of eigenvectors displayed how each trajectory explored the conformational space (Figure 4.2). The difference in the initial velocities can sometimes allow the protein to get rid of the initial biases and let the protein sample diverse conformational states. The two simulations Sim1 and Sim2, started from the crystal structure, explored very different regions in the vectors space of PCA1 and PCA2 before settling in a small region. This indicated that the two simulations sampled various conformations to get to different equilibrium states.

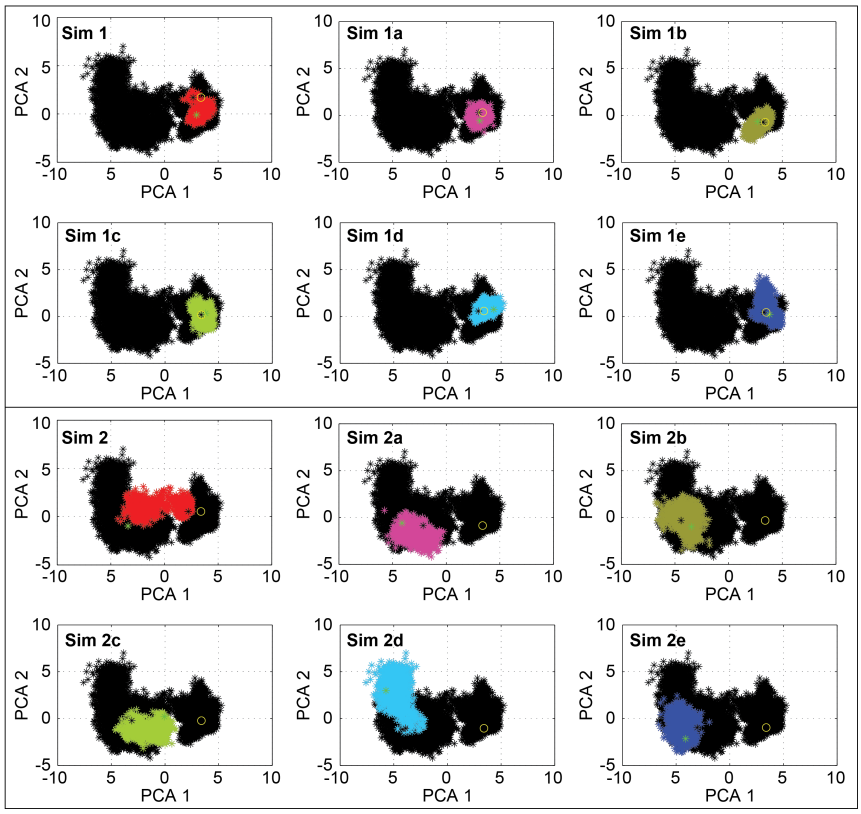


Figure 4.2: Projections of the individual wildtype trajectories on the first two eigenvectors obtained from the PCA of the concatenated set. For each graph the black dots represent the concatenated set and the colored dots are the individual trajectories as labelled on the plot.

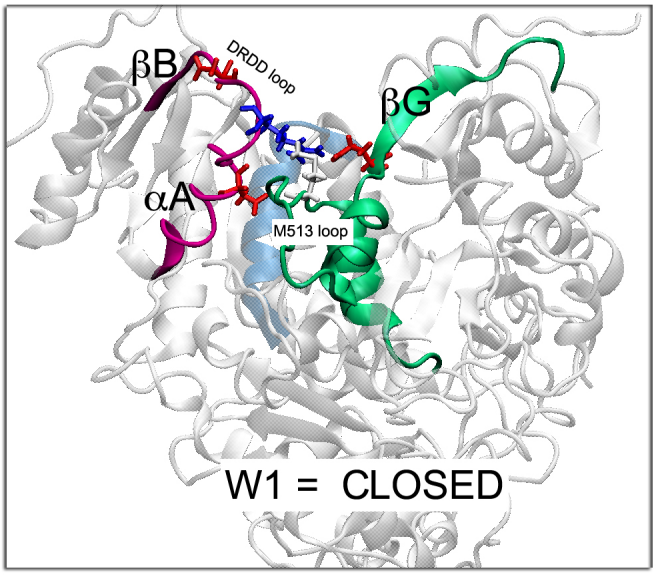
The simulations from their end points, Sim1a-e and Sim2a-e, were mostly seen to sample around the same region indicating that they had reached a local minima.

This analysis revealed that a single simulation for around 50 ns or more allowed the protein to explore a conformational region centered around only one type of population but by starting several trajectories with different velocities many different kinds of populations could be witnessed.

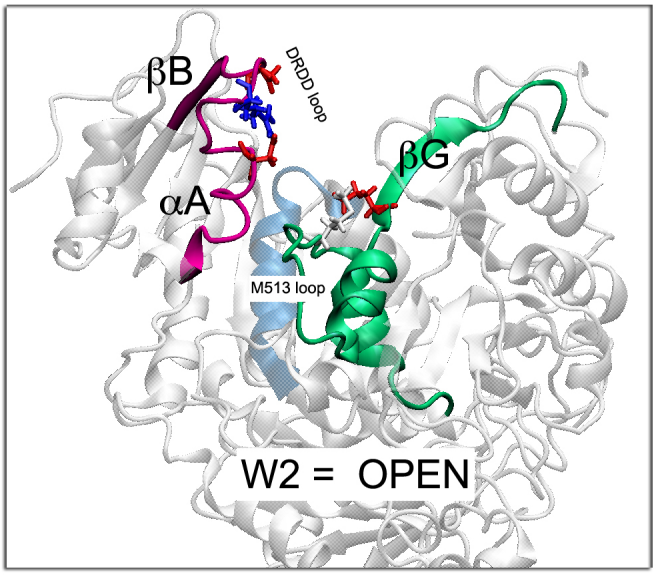
Two Major Conformations in the Wildtype Simulations

The projection of all the wildtype trajectories on the first mode displayed two peaks located away from each other (Figure 4.1a and 4.5). Separation of the trajectory in these peaks indicated that there were two major types populations in the protein, each representing a conformational state, and the largest motion of the protein took place in moving between the two states. The representative structures were extracted at each peak (Figure 4.3(a) and 4.3(b)) and they showed that the main difference between these two populations was the distance between two loops of the protein, the DRDD loop and the M513 loop. DRDD is the inter–loop between αA and βB and M513 loop lies between αG and αH. The conformation representing the left peak was named as W2 and that representing the right peak as W1. Structural analysis showed in W1, the DRDD made local interactions with the residues of the M513 loop and βG strand, or in other words the components of the calcium binding site were located close to each other. W2, on the other hand, had the DRDD loop located far from the M513 loop taking the important calcium binding partner away from the other components.

The PCA analysis indicated that the distance between the two loops might be the major difference between the two populations. In order to find the extent of this movement in the 3-dimensonal space, the distance between these loops was calculated in all snapshots. The distribution of this distance yielded two peaks; one at 6.5 Å and the other 13.5 Å (Figure 4.5). On overlaying the distance distribution on the projection values obtained before, it was found that the 6.5 Å peak coincided with W1 population and the 13.5 Å peak coincided with W2. Thus, it showed that in moving from W1 to W2 state, the structural elements of the protein traveled approximately 7 Å distance in space which reflected the dynamic character of the wildtype protein.



(a) W1, Closed conformation



(b) W2, Open conformation

Figure 4.3: The structures of the two populations W1 and W2 as identified from the PCA analysis. The location of the DRDD loop, M513 loop and the structural elements surrounding these loops have been labelled. Note the difference in the distance between the DRDD loop and M513 loop in the both the states.

Unstable State of the Wildtype Protein

The projections on the PCA modes exposed another interesting feature in the wildtype protein dynamics. Contour plots of the projections on PCA1 and PC2 showed that almost 90% of the trajectory settled in regions lying outside the values 0 and 1 on PCA1 (Figure 4.4). This conformational region was named as W3 and the snapshots

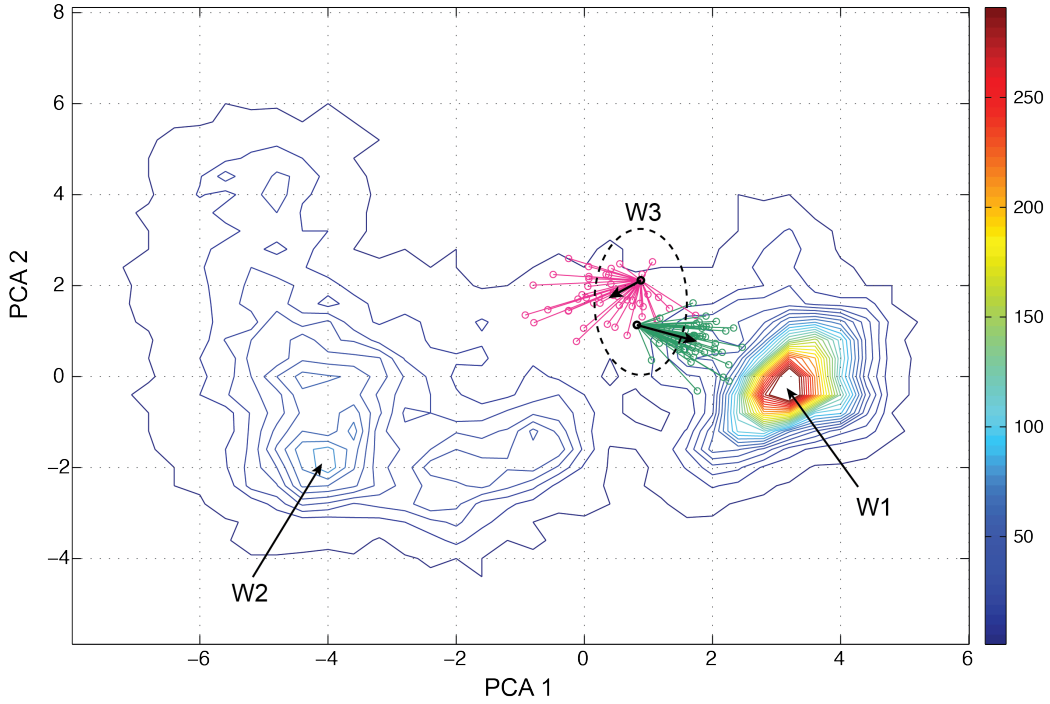


Figure 4.4: Representation of the unstable region in the wildtype trajectory. Contour plot of the projection of wildtype trajectory on PCA1 and PCA2 of wildtype basis set show an unstable region W3 between PCA1 values 0 and 1 and PCA2 values 1 and 2. 50 short trajectories were started from two structures lying in this region. The path of these trajectories are shown by pink lines and green lines and the resultant direction of all trajectories is shown by the black arrow. Both the sets show a resultant movement towards the nearest peak indicating that conformational space W3 was were very unstable.

lying in this region were selected for further analysis. Two structures were randomly picked from this region and were used to run several short simulations (100 ps long), all with different initial velocities. The projection of each of these short trajectories, when plotted on Figure 4.4 showed that none of them stayed in W3. Even in the short span of 100 ps, each trajectory moved away from W3 and were put on a path

which lead to the closest peak (represented by arrows in Figure 4.4). This analysis confirmed that the conformation of the protein in W3 region was highly unstable.

4.3.3 Characterization of the Mutant Protein

Next, the structural and dynamical features of the mutant protein were explored. The principal modes of the combined mutant trajectories were obtained in the same way as seen in the wildtype case. The analysis showed that the amplitude of the most significant motion of the mutant protein had reduced considerably. The magnitude of the eigenvalue of the first mode of D369G protein became less than half of that seen in wildtype simulations (from 12.31 nm^2 to 5.86 nm^2). Also, the largest motion of the mutant protein no longer represented the transition between conformations W1 and W2. Instead the dynamics in this mode was represented by the movement of a loop lying between αB and βC.

Mutant trajectory between W1 and W2

The wildtype simulations showed above that the protein was very dynamic and the large motion was represented by the DRDD loop moving relative to the M513 loop. In order to find out what happened to this motion in the mutant protein, the mutant trajectory was projected on the same basis vectors. The projection on the first mode showed a single peak at a conformational region lying between the wildtype peaks W1 and W2 (Figure 4.5). This revealed that in the mutant protein the two loops, DRDD and M513, no longer moved away from each other. The distance between these two loops in the mutant trajectory (represented by blue dots in Figure 4.5) showed a single peak at 7 Å indicating that the mutant protein was not that dynamic.

Further, these projections showed that the mutant trajectory populated W3, a region of the wildtype conformational space found to be extremely unstable. Thus, it looked that the mutation changed the backbone conformation of the protein in such a way that it became stable in W3 region and stayed there. The next set of analysis have been focussed on identifying the structural differences between the two proteins which could lead each protein to exhibit the aforementioned behavior.

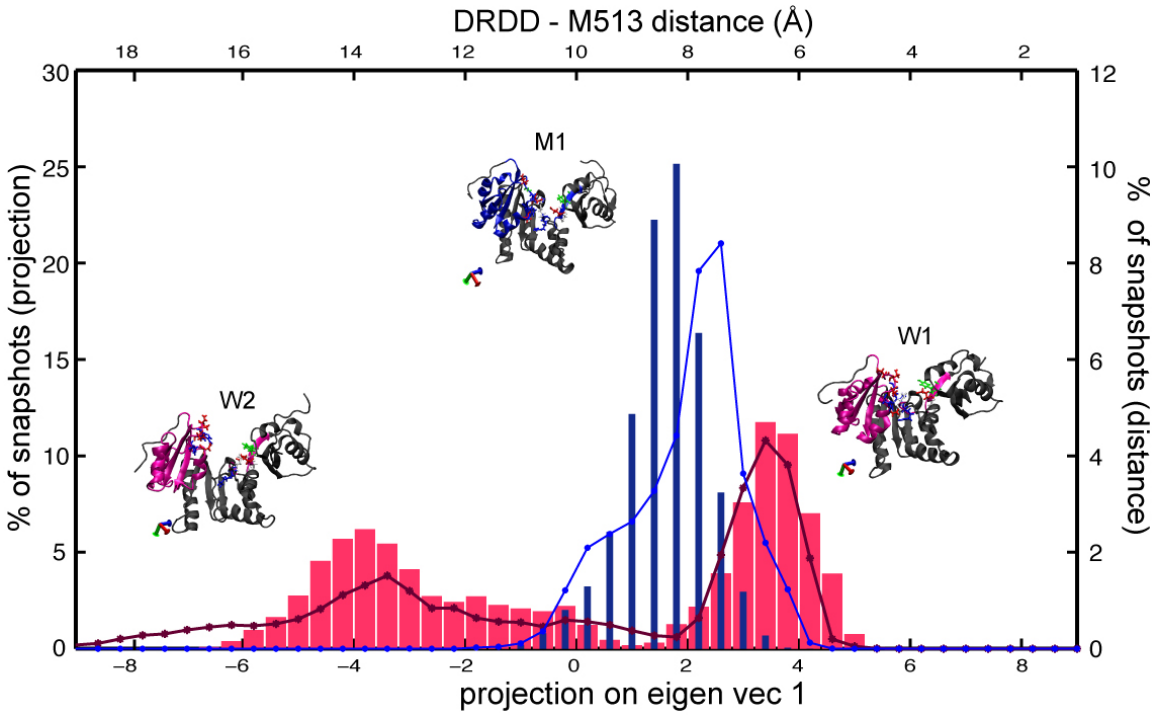


Figure 4.5: The projection of the wildtype and mutant trajectories on the PCA1 of wildtype basis set. Wildtype trajectory projection are in pink bars and the mutant trajectory projection in blue bars. The structures W1, W2 and M1 represent the conformations of the peak closest to them, W1 for the right peak in wildtype projections, W2 for left peak in wildtype projections and M1 for the single peak in mutant projections. Red and blue lines overlaying the projection values are the distribution of the distance between the DRDD loop and M513 loop in the wildtype and mutant trajectories respectively. The top axis is for the distance plot and the bottom one is for the projections.

Secondary Structural Changes

To investigate the change in the backbone conformation of the protein, the distribution of the backbone angles ϕ and ψ were calculated over the entire trajectory. The change from an Asp to a Gly is known to cause a significant change in the backbone of the protein as it results in removal of the bulky Asp side chain. The $\phi - \psi$ angle distributions showed that the change to Gly indeed changed the secondary structure of the residue 369 (Figure 4.7). The trajectory started spanning the positive region of the Ramachandran plot which was expected for the Gly residue. The interesting observation was for the neighboring residues, Asp367 and Arg368 (Figure 4.6). Both

of them saw a considerable shift in the Ramachandran plot upon mutation at location 369. Asp367 though stayed in the same quadrant but showed peaks at different values of ϕ and ψ. Arg368 moved out of the top left region and now showed much lower values of ψ. Thus, the analysis revealed that the mutation changed the local backbone conformation of the DRDD loop.

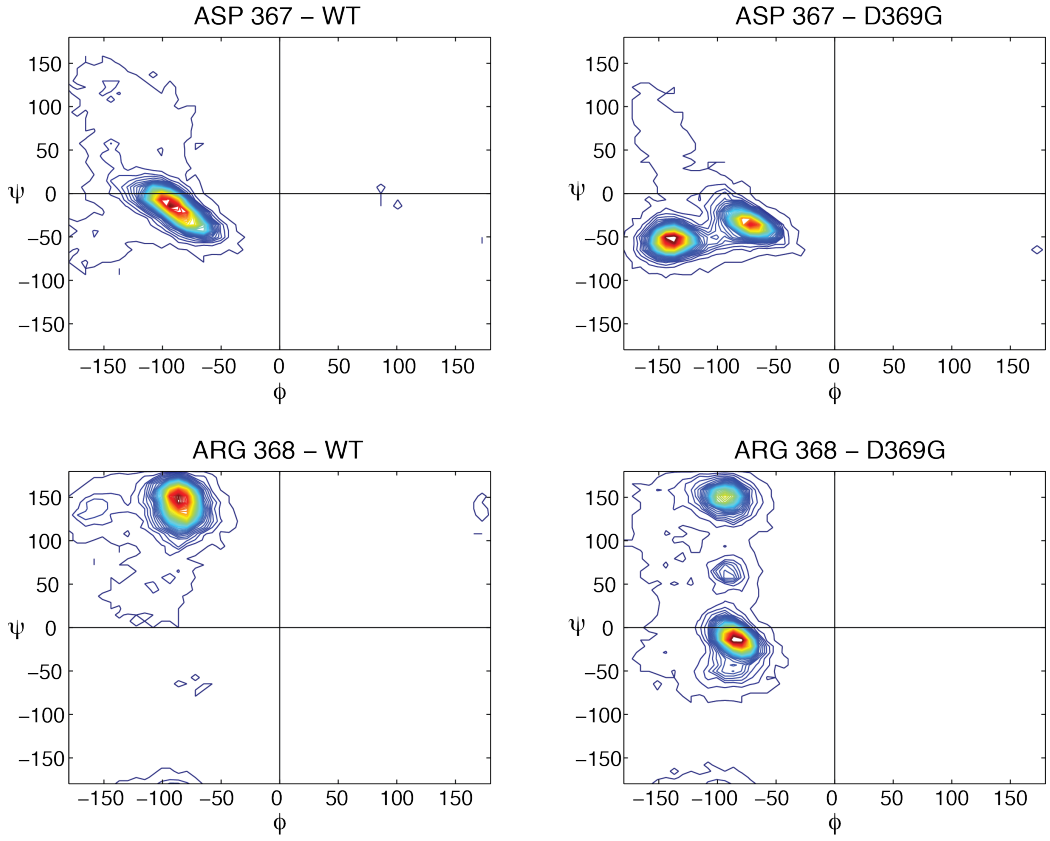


Figure 4.6: Comparison of the $\phi - \psi$ angle distributions for the DRDD loop residues, 367 and 368 in wildtype and mutant simulations. The plots show that both the residues experience a considerable change in their backbone angles after the mutation

Residue-Residue Interaction

The change in the DRDD loop conformation suggested that the local interactions of the loop residues might experience a significant change. To find that out, statistics on the salt bridge and hydrogen-bond formation was collected for all trajectories. The interactions were indeed very different in the wildtype and the mutant proteins. In general the mutant protein appeared as a closely packed structure with more number

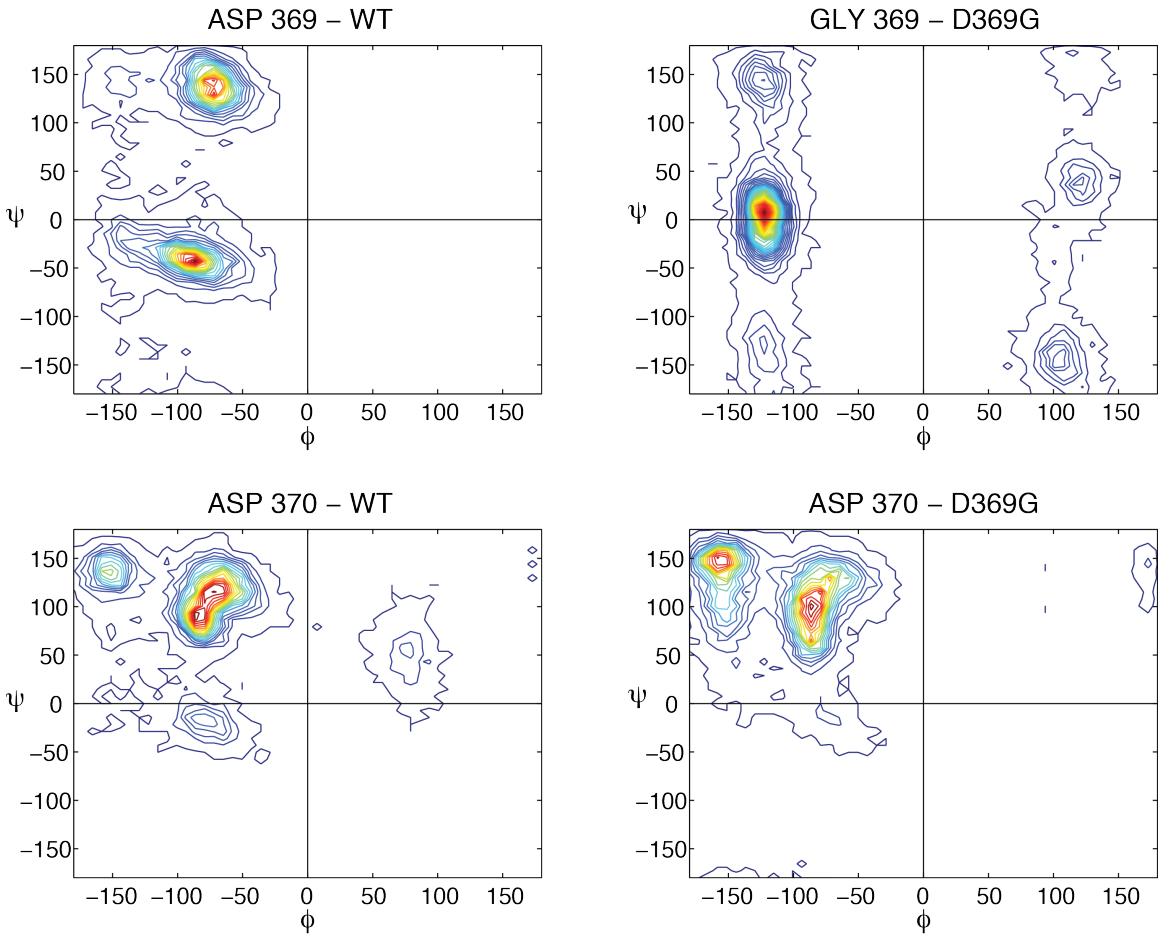
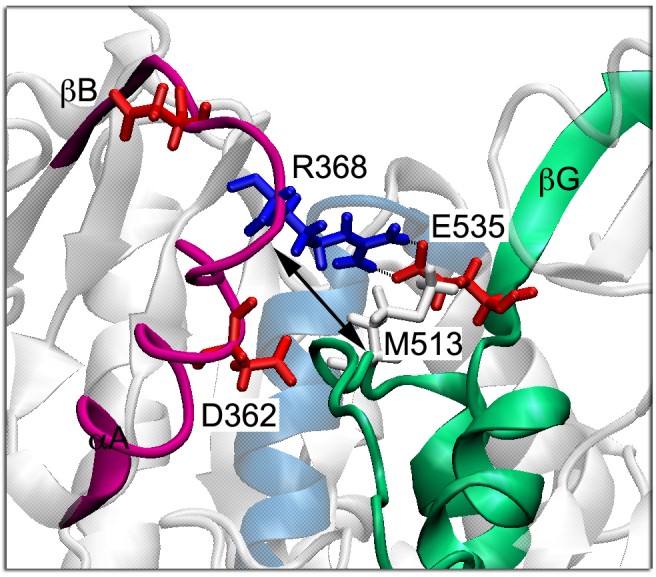
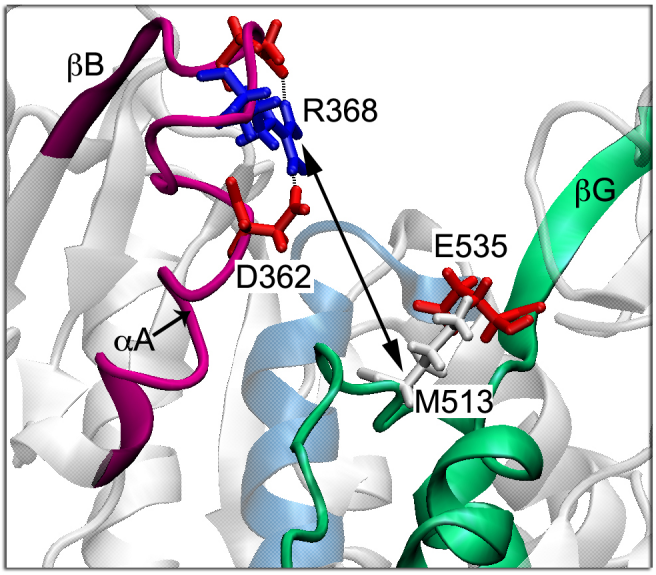


Figure 4.7: Comparison of the $\phi - \psi$ angle distributions for the DRDD loop residues, 369 and 370 in wildtype and mutant simulations. The changes in the backbone angles from Asp to Gly at location 369 are as expected. The residue 370 does not experience any change upon mutation at location 369.

of interactions between the αA helix and the RCK2 helices. The wildtype protein showed two sets of interactions, one corresponding to W1 and the other to W2. W1 showed interactions connecting DRDD loop to structural elements in farther domains such as, the M513 loop and βG in RCK1 domain and several helices in RCK2 domain. The W2 state was seen to contain less domain-domain interactions. The DRDD loop was contained in itself, making salt bridges with partners located 2-3 residues upstream or downstream of itself. The interactions important in determining the DRDD loop behavior have been listed in Table 4.1 with their percentage of formation in each protein. The % values revealed that the residue Arg368, in the DRDD loop, completely changed its interacting partners from the wildtype to the mutant protein.



(a) W1, Closed conformation



(b) W2, Open conformation

Figure 4.8: The structures of the two populations W1 and W2 as identified from the PCA analysis. The salt bridge formation of residue Arg368 is different in both. W1 shows Arg368 interacting with Glu535 and W2 shows its interaction with Asp362 and Asp340.

In the wildtype protein, it was seen to interact with Asp362, Asp 340 and Glu535 but in the mutant protein it almost always interacted with Asp367.

Interactions	wildtype	D369G	Difference
Dominant in wildtype			
ARG368-GLU535	46%	7%	+39%
ARG368-ASP362	38%	2%	+36%
ARG368-ASP340	20%	0.01%	+20%
Dominant in Mutant			
ARG368-ASP367	3%	90%	-87%
ASP362-SER922	0.6%	80%	-80%
ASP362-SER925	1.0%	79%	-80%

Table 4.1: Residue-residue interactions in wildtype and D369G proteins.

Molecular Mechanism from W1 to W2

The results shown above indicated that Arg368 might have an important role to play in generating the two populations of the wildtype protein. The Arg368–Asp 340 interaction was found to be anti-correlated with the Arg368–Glu535 interaction in the wildtype protein. The former was formed in the W2 state (Figure 4.8(b)) and the latter in the W1 state (Figure 4.8(a)).

The intermediates between the W1 and the W2 state, showed that Arg368 gradually changed its interactions in going from one state to the other. Figure 4.9 shows these intermediates. From these structures it appears that the following mechanism might be taking place. In the W1 state the salt bridge between Arg368 and Glu535 was keeping the two loops, the DRDD loop and the M513 loop together which probably gave rise to the W1 population. When this interaction was lost, Arg368 searched for other negatively charged residues and found Asp362. This salt bridge pulled Asp362 towards the Arg368 and brought secondary structural changes in αA helix which resulted in breaking the interactions of αA with the rest of the protein. This process continued until the domain was isolated and gave rise to the W2 population.

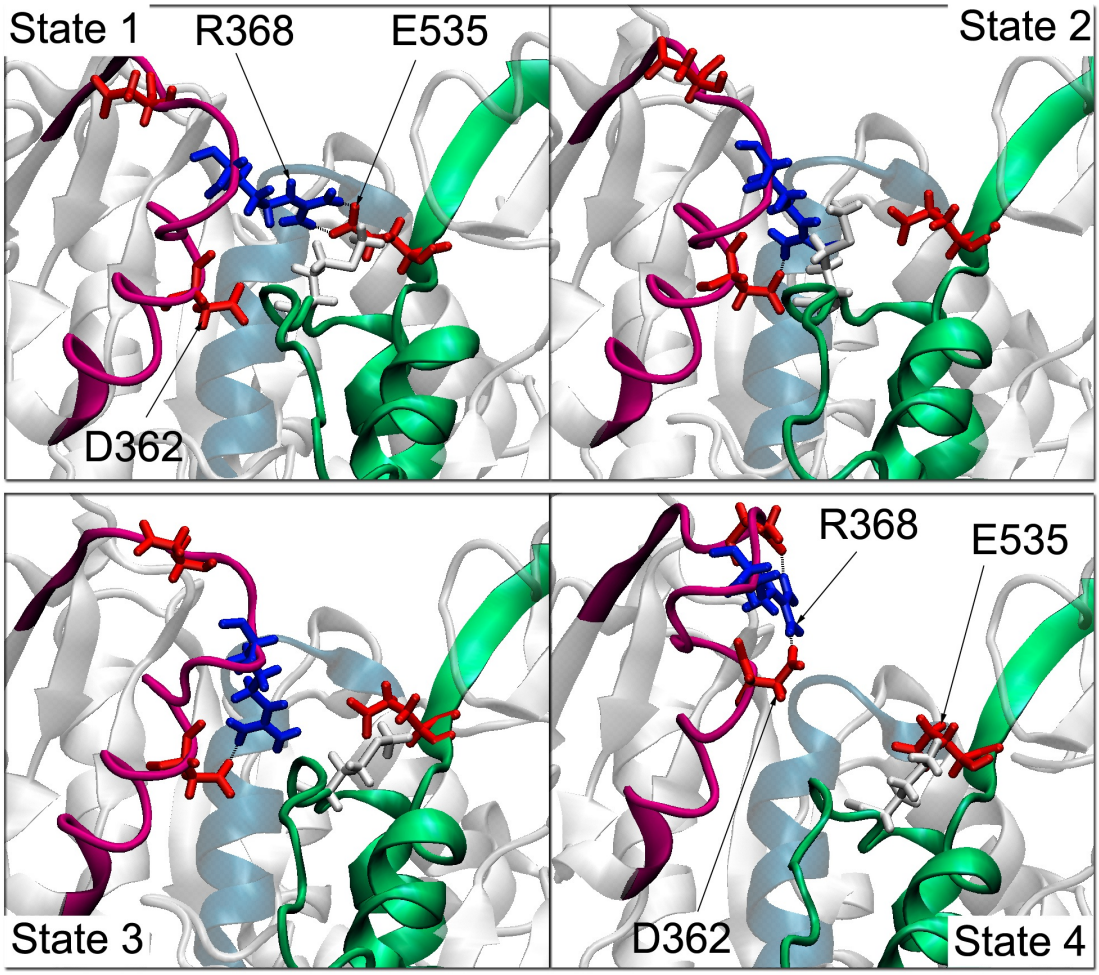


Figure 4.9: Molecular mechanism of the transition of the wildtype trajectory from the closed state to the open state. Arg368 forms a salt bridge with Glu365 in the W1 state (Step 1). This bond is broken and Arg starts interacting with Asp362 (Step 2). The interaction pulls Asp362 towards Arg, thereby pulling αA away from the rest of the protein (Step 3). The pull results in complete isolation of the domain and gives rise to the W2 state (Step 4).

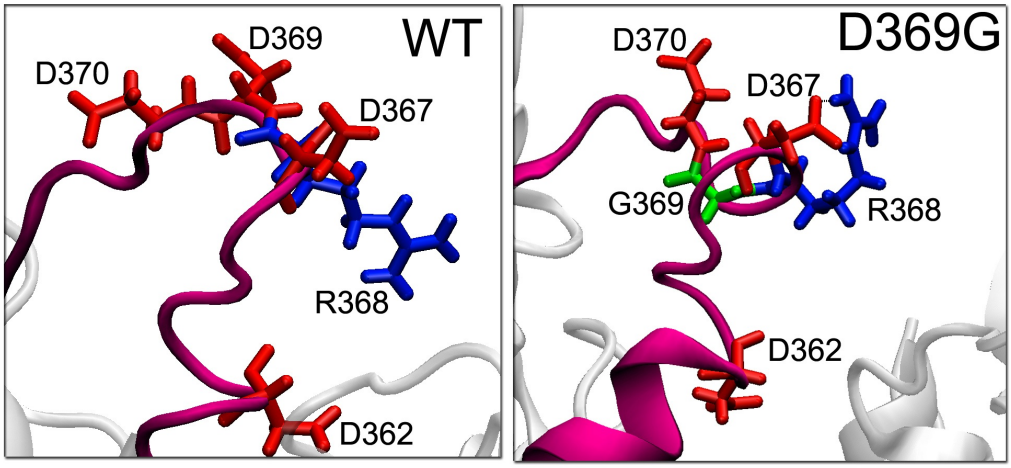


Figure 4.10: The D369G mutation changes the conformation of the DRDD loop. In the wildtype protein (left) all four amino acids are spread apart and interact with residues in other domains. In the mutant protein, the presence of Gly369 inserts a kink, which allows the DRDD residues to interact among themselves. Arg368 is seen to coordinate with Asp367.

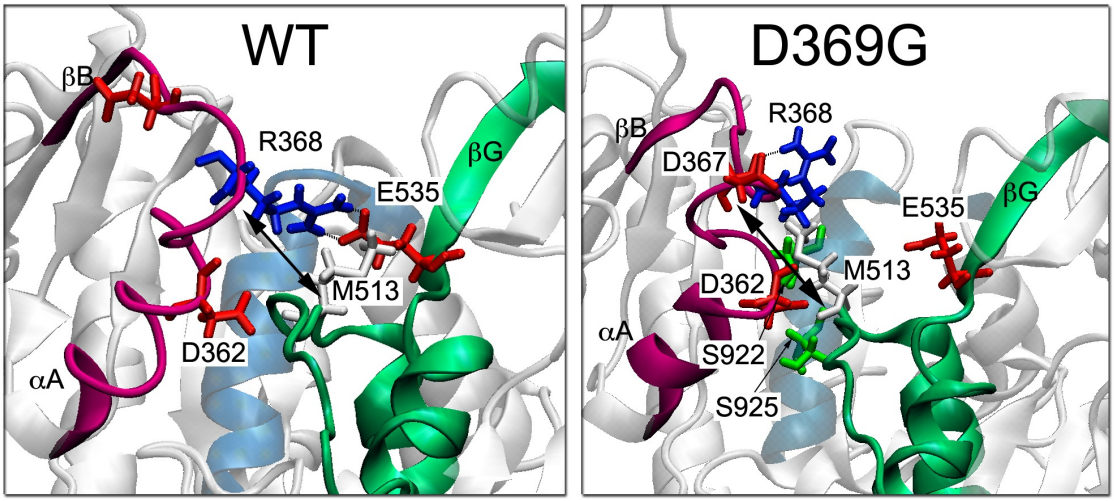


Figure 4.11: The difference between the reside-residue interactions in the wildtype closed state (left) and the mutant state (right). Due to the change in the backbone caused by insertion of Gly369 in the mutant protein, Arg368 forms a salt bridge with Asp367. This leaves Asp362 free to interact with other residues and it is seen to stay close to the RCK2 residues, specifically Ser922 and Ser925 of αS helix. These interactions allow the protein to initiate more connections between αA, αS and αG which help in improving the stability of the conformation.

Effect of Mutation

In the mutant protein, this mechanism was not observed. The presence of Gly at position 369 here was seen to introduce structural changes in the DRDD loop which allowed Arg368 to interact with residues of the loop itself, Asp367 (Figure 4.10). This interaction engaged the Arg almost always (90%) and this probably did not leave it to interact with Asp362. The result was seen in the orientation of the αA helix. In the wildtype case it was being pulled towards the DRDD loop, away from the rest of the protein, but in the mutant case it was seen to stay in close proximity to these domains (Figure 4.11). Hydrogen bond interactions showed that in the mutant case, Asp362 in αA helix formed strong interactions with Ser922 and Ser925 of the αS helix in RCK2 (80%). A careful observation of the structure and interactions showed that the protein now attained a packed conformation with a higher frequency of non-bonded interactions than observed in the wildtype case. Thus, we can conclude that the structural change introduced by glycine in the DRDD loop, drives the RCK1 domain to a rigid conformation which is much more stable.

4.4 Discussion and Conclusions

D369G mutation, in the BK_{Ca} channel was known to make the channel hyperactive and more sensitive to calcium ion. In this chapter, we investigated the structural effect of this mutation and identified the molecular mechanism which is modified by the mutation. The mutation site is located in an important segment of the BK_{Ca} channel structure, the DRDD loop. This loop, along with M513 loop and βG have been thought to enclose a Ca^{2+} binding site, and mutagenesis studies have identified residues Asp367, Met513, Arg514 and Glu535 as being important for the calcium related function of the channel. We performed molecular dynamics simulation of the entire subunit of the channel and found that the protein exhibited large scale dynamics in the aforementioned binding site region. The DRDD loop was seen to move as far as 10 Å away from the M513 loop and the βG strand. When we mutated the residue Asp369 to Gly, we observed a significant loss in the magnitude of this dynamics. The findings from our simulations can be summarized as following:

- PCA showed that the wildtype protein was very dynamic exhibiting two major conformations, W1 state and W2 state.
- The difference between the two conformations was the change in distance between the DRDD loop and the M513 loop from 6.5 Å to 13.5 Å.
- Hydrogen bond and salt bridge interactions revealed that this dynamics was governed by the change in interactions of Arg368, in the DRDD loop. In the closed state Arg368 interacted with Glu535 of the βG strand and kept the DRDD loop in close proximity to the rest of the protein. In the open state it interacted with Asp362 and Asp340, which left the DRDD isolated and away from the remaining domains.
- Mutation of Asp369 to Gly in the D369G mutation changed the conformation of the DRDD loop. Secondary structural analysis showed that mutation changed the protein backbone at the site of the mutation and also around it.
- This completely altered the interaction partners of Arg368. Hydrogen bond and salt bridge statistics showed that Arg368 which was interacting with Glu535, Asp362 and Asp340 in the wildtype protein now interacted with Asp367 and stayed there almost all the time.
- This drove the protein to a new conformational state which was between the closed and the open wildtype states. The distance between the DRDD loop and the M513 loop in this conformation was seen to peak at 7 Å .
- The protein became less dynamic and remained stable in this conformation.

We have used molecular dynamics simulations to investigate the molecular mechanisms in the BK_{Ca} channel protein and the results suggest that the dynamics observed in the wildtype MD simulations contain functionally relevant conformational fluctuations. The lack of Ca^{2+} in the simulations prevents us from drawing detailed conclusions about this mechanism but by looking at the effect of the D369G mutation we can get some important insights.

Through our simulations we find that the Ca^{2+} binding site in the RCK1 protein, D367/E535 site is very dynamic in the wildtype protein. The dynamics is mostly

caused by the movement of the DRDD loop towards and away from the rest of the protein. Previous studies have shown that Asp367 in the DRDD loop binds to the Ca^{2+} ion. This would mean that when the Ca^{2+} binds at this site, it will hold the DRDD loop in a fixed conformation and will thus reduce the overall dynamics. Interestingly, the MD simulations of the D369G mutation show exactly this phenomenon, except for the fact that the reduction in dynamics is achieved by the mutation and not by binding of the Ca^{2+} ion.

Thus, it is very suggestive that the mutation might be mimicking the effect of calcium binding. This was certainly seen in the experimental results of this mutation. In BK_{Ca} channels, the role of Ca^{2+} ions is in upregulating the activation machinery (Meera et al., 1996). Single channel recordings showed that increase in the intracellular Ca^{2+} concentration caused an increase in the open probability of the channel. Similar experiments on the D369G mutation showed that the mutant channel showed similar upregulation at a much lower Ca^{2+} concentration, indicating that the mutation mimics the Ca^{2+} binding effect (Du et al., 2005).

Chapter 5

Multisite Cation Model and its Applications

5.1 Introduction

After identifying the effect of the mutation on the Ca^{2+} free form of the BK_{Ca} channel protein, the research in this thesis led to the studying the effect of the mutation on the Ca^{2+} bound structure. A combination of implicit solvent and explicit solvent techniques were employed to study the effect of Ca^{2+} binding and extra care was taken to ensure that the ion was assigned accurate parameters. Careful reparametrization of the calcium ion was performed by matching the ion-solvation energies and the values were used for calculating the interaction of the Ca^{2+} ion with the channel protein. These simulations, however, posed several challenges. Many cases were found where the Ca^{2+} ion was either too tightly bound to the protein such that it could coordinate with a single atom or else it unphysical conformations of the protein were prompted. Researchers earlier have faced similar problems and over the past few decades several improvements have been suggested such as reparametrization of the ion with the target protein, using specialized potentials, integrating ab-intio quantum mechanical calculations in the molecular mechanics calculations or using polarizable force fields, but most of these methods are computationally expensive.

In all of these methods, the charge on the cation is fixed and its value is obtained by matching the solvation energy of the ion in bulk water. In bulk water the atoms coordinating with the ion are free to move and hence when running simulations, the atoms are able to quickly arrange around the ion reach. Simulations of proteins,

however, require longer time for equilibration, which is usually not enough for the equilibrated arrangement of the coordinating atoms around the ion. And, by fixing the charge in the center of the ion model, there are not many dials left to turn.

In this chapter, a new model of the cation is proposed which does not carry a fixed charge in the center of the ion. The model, named as the multisite cation model, splits the charge of the cation in smaller values, in the direction of the coordinating atoms. Thus, the force of interaction between the coordinating atoms and the ion gets a direction and it helps in guiding them in accurate geometry around the ion. The model parameterized in bulk water can be easily transferred to proteins because now the small charge centers have some amount of flexibility to identify the accurate coordinating partner and reproduce accurate microscopic interactions.

The chapter presents the methods used for development and parametrization of the multisite cation model and the molecular dynamics simulations of this model in water and with ion-binding proteins. The simulations showed that the model provides a great deal of accuracy in the ion–protein interactions and successfully reproduce accurate structural and thermodynamics properties of the coordinating ligands.

5.2 Cations and Coordination Geometry

Cations are known to interact with the negatively charged atoms by forming coordinate bonds. A coordinate bond is unique, as it has the characteristics of both an ionic and a covalent bond. Here, both shared electrons between the atoms are provided by the ligand, to the empty shell of the metal ion The metal ion interacts with more than one ligand and these interacting atoms arrange around the metal ion in a particular geometry. Pauling developed a set of principles which governed the structures of the complex ionic crystals. These principles later came to be known as Pauling Rules and have been described in his article "The principles determining the structure of complex ionic crystals" (Pauling, 1929). The rules were as follows:

- **Rule 1**: A coordination polyhedron of anions is formed around each cation, where the cation-anion distance is determined by the sum of the ionic radii, and

the coordination number of the polyhedron is determined by the cation/anion radius ratio (Rc:Ra).

- **Rule 2**: The electrostatic valency principle. The strength of an ionic (electrostatic) bond (electrostatic valency e.v.) between a cation and an anion is equal to the charge of the ion (z) divided by its coordination number (n): e.v. = z/n. In a stable (neutral) structure, a charge balance results between the cation and its polyhedral anions with which it is bonded.
- **Rule 3**: Sharing of faces or edges is unstable.
- **Rule 4**: In structures with different types of cations, those cations with high valency and small coordination number tend not to share polyhedra with each other; when they do, polyhedra are deformed to accommodate cation repulsion.

Thus the cations with coordination number 4 were seen to have a tetrahedral geometry, 6 had an octahedral geometry, 7 was a pentagonal bipyramid and so on. The ideal cases were seen in crystals, but even with the proteins, the geometry of the ligands around the cations was more or less conserved. The crystal structures of proteins showed that the metal binding site had the coordinating atoms arranged in a particular geometry. The Mg^{2+} binding site was seen to have an octahedral arrangement, the Ca^{2+} binding site was mostly pentagonal bipyramid, though octahedral cases were also seen, the zinc site was mostly tetrahedral, and so on (Falke et al., 1994; Strynadka and James, 1989).

5.3 Current Limitations of Representing Cations in Molecular Dynamics Simulations

Cations play a major role in proteins. They are instrumental in regulating the vital physiological functions of the body such as folding of proteins and nucleic acids, enzyme catalysis, numerous cell signaling processes and certain catalytic and regulatory activities of proteins and enzymes. To probe and explore the accurate structural and functional roles of ions in the biological phenomena, it is important to reproduce accurate and physically realistic descriptions of ion-interactions with the biomolecules.

It can be challenging for experimentalists to study the metal mediated mechanisms and hence molecular dynamics can be a great tool for studying such mechanisms. However, accurate computational models for the ions are required. Researchers have produced a number or non-polarizable and polarizable models of the ions but it has been very challenging to strike a balance between the accuracy of the model and the computational expense.

The parameters have been obtained by accurately matching the solvation free energy of ions in water and these values are being regularly used in other cases as well (Aaqvist, 1990). Recently, newer models have been developed which take account of the many body affects of the ionic systems such as induced electronic polarization (Joung and Cheatham, 2008) and efforts are also being directed to using the polarizable force fields for such systems (Yu et al., 2010) or *abinitio* quantum mechanical integration in the molecular mechanics potential (Asthagiri et al., 2010).

For simulations of proteins it is necessary to attain proper statistical averaging which require longer simulations allowing the systems to sample over many configurations. The ion-water models parameterized in water are being used in proteins but this introduces inaccuracies making it longer for the system to reach equilibrium. This is because in proteins the coordinating atoms are not as free to move as in water and it takes much longer for them to obtain an accurate coordinating geometry. Also, the currently used models are fixed charge models where it becomes challenging to find a general value with respect to the different type of protein binding sites. Researchers have come up with several solutions to get rid of these inaccuracies by changing the parameters of the coordinating atoms, or by putting external constraints between the ion and the coordinating atoms (Cates et al., 2002; Allouche et al., 1999) but these have to be specifically defined for every system.

In this chapter, a new computational model of the cation is being introduced which does not contain a fixed charge in the center of the cation. Instead the net charge is split into n dummy charge centers in the direction of the coordinating atoms. The model called the multisite cation model has been derived from the structures of the quantum optimized solvation shells of cations. The following sections go over the methodology used in the development and parameterization of this model. Once selected, the parameters are used in simulations of ion with water and with prototypical

ion binding proteins. The last section shows how the model can be successfully used in reproducing thermodynamic properties of complex mechanisms such as comparative selection of ions by proteins.

5.4 Design of Multisite Cation Model

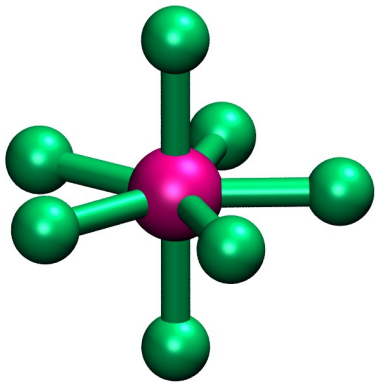


Figure 5.1: The calcium multisite model ($n = 7$)

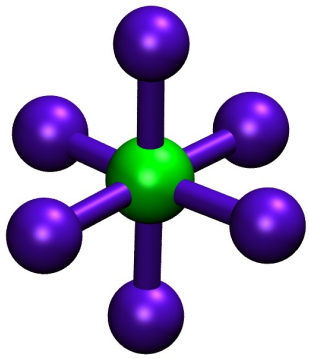


Figure 5.2: The magnesium multisite model ($n = 6$)

The multisite cation model was developed by replacing the spherical representation of the cation with a molecule containing 1 van der Waals center (M) covalently bonded to n dummy charge centers (D_i), where n is the coordination number of the cation. The divalent ions Ca^{2+} and Mg^{2+} have been used in this study to develop their respective multisite models. The geometry of the charge centers was derived from the coordination geometry of the cation. The multisite cation models were developed as polygon of n charge centers enclosing a center atom M at the center. Mg^{2+} has a smaller ion radius and it has been observed to favor a strict sixfold, octahedral coordination geometry (Falke et al., 1994). Ca^{2+} has a bigger radius and it has been seen to have 6 to 8 coordinating atoms. The geometry for most of the Ca^{2+} coordinating atoms has been observed as a pentagonal bipyramid, coordinating with 7 atoms (Strynadka and James, 1989). Thus the polygon for Mg^{2+} ion was developed as an octahedral geometry (Figure 5.2) and that for the Ca^{2+} ion got a pentagonal bipyramidal geometry (Figure 5.1).

Each edge of the regular polyhedron was represented as a covalent bond with 0.9 Å equilibrium bond length and 540 kcal/mol/Å^2 of bond stretching force constant. The covalent description and length of this bond matched with the earlier description of similar models (Aqvist and Warshel, 1990, 1992; Pang et al., 2000; Park et al., 2006).

	k	R_o/θ_o
Bonds	540 (kcal/mol/Å^2)	0.900 Å
Angles	55 (kcal/mol/rad^2)	72, 144, 90 and 180 (°)

Table 5.1: Bonded parameters of the multisite cations.

The parameters for all the non–bonded terms in the molecular mechanics force field were obtained systematically such that the experimentally derived energies were reproduced accurately. Previously, the parameterization for an ion model required matching the ion-solvation energy to the experimentally obtained values. In the present case, since there were more number of variables, a systematic refinement procedure was applied which involved matching three sets of energy values and the final values were selected such that most of the criteria were satisfied.

5.4.1 Parameterization of the Model

The cation is involved is mainly two groups of non–bonded interactions, electrostatics and van der Waals. The electrostatic interactions are calculated as a sum of interactions between pair of point charges, using Coulomb's law:

$$E_{elec} = \sum_{i}^{N_A} \sum_{j}^{N_B} \frac{q_i q_j}{4\pi\epsilon_o r_{ij}} \tag{5.1}$$

The van der Waals interactions are calculated by the Lennard–Jones potential:

$$U_{tot} = \frac{A_i * A_j}{r_{ij}^{12}} - \frac{B_i * B_j}{r_{ij}^{6}} \tag{5.2}$$

For the multisite cation model, the required parameters were; the charge on each atom, q_i and the Lennard–Jones constant A_i and B_i. The total charge of the cation was distributed on each atom (M and n dummies D_i) such it matched the charge of the ion, $q(M)+nq(D_i) = q(cation)$. The n dummy atoms were assumed to be similar

atom types, which meant that each of them carried the same charge and had the same Lennard–Jones constants. To identify the unknown variables: $q(M)$, $q(D)$, A_M, B_M, A_D, B_D, the following energies were matched with their experimental values:

1. Fitting the cation-oxygen **Lennard–Jones potential** to the potential obtained from using the standard parameters of cations
2. Matching the **single point energy** of the cation-water solvation structure, from molecular mechanics to the corresponding value from the quantum calculations
3. Matching the **solvation free energy** obtained from thermodynamic integration calculations to the experimentally obtained value

With these three criteria, there was not a unique set of parameters which was obtained. The energies were calculated by varying all the variable systematically and then those values were selected which showed the best correlation with all experimental energy values. The aforementioned procedure was performed individually for Ca^{2+} and Mg^{2+} ions.

1. Fitting the Lennard–Jones Potential

To calculate the Lennard Jones potential for the multisite cation interaction with oxygen the following equation was used 5.3:

$$U_{tot} = \frac{A_M * A_o}{r_M^{12}} - \frac{B_M * B_o}{r_M^6} + \sum_d^n \{\frac{A_d * A_o}{r_d^{12}} - \frac{B_d * B_o}{r_d^6}\} \tag{5.3}$$

This equation accounted for the interaction of the ion with the oxygen in a single orientation. Since each dummy atom had a different distance with the oxygen atom, the energy value varied when the ion rotated. The spread of the energy profile at a particular combination of A and B (seen in Figure 5.3) demonstrated the variations in energy upon that rotation. This variation was incorporated by following Euler's rotation theorem and an average energy value was calculated for 1000 rotations. These average values were obtained for different combinations of A_M, B_M, A_D, B_D and each were fit to the standard parameters of cation–oxygen interaction by using Nelder-Mead simplex algorithm. The standard parameters were obtained from the Aqvist's paper (A_{Ca} = 264.1; B_{Ca} = 18.82; A_O = 762.89; B_O = 24.39; A_{Mg} = 37.0; B_{Mg} = 8.32 (Aaqvist, 1990); units A=((kcal/mol)Å^{12}) and B=((kcal/mol)Å^6)). The combination

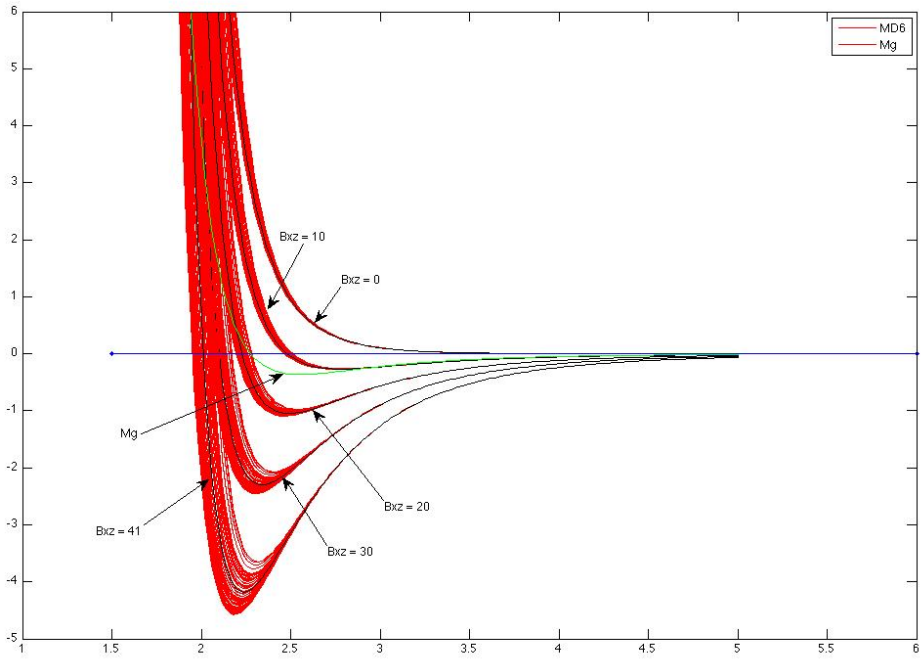


Figure 5.3: Cation–water Lennard–Jones interaction energy obtained for the multi-site cation with different combinations of A_M, B_M, A_D, B_D. The spread at each combination is obtained by 1000 Euler rotations of the ion.

which gave the best fit with the standard parameters were recorded to compare with the values which will be obtained from the subsequent energy evaluations.

2. Single Point Energy of the Cation–Water Complex

Complexes of the cation, $[Ca(H_2O)_6]^{2+}$ with the first solvation shell oxygen atoms and $[Mg(H_2O)_6]^{2+}(H_2O)_{12}$ with the first and second solvation shell oxygen atoms were used to find the single point energy through quantum mechanical calculations using GAUSSIAN. These calculations were divided into two steps: first a geometry optimization was performed to get the optimized structure of the complex and then the single point energy was evaluated. B3LYP method and 6-311+G(2d,2p) basis set were used for both steps. 6-311+G(2d,2p) basis set has two sets of polarization functions on all atoms and diffuse functions which are important for interactions with

oxygen in water containing systems. A smaller basis set LANL2DZ was also tried for optimizations but the former gave better energies.

The optimized structures obtained for the ($[Ca(H_2O)_6]^{2+}$) and the $[Mg(H_2O)_6]^{2+}(H_2O)_{12}$ complexes are shown in Figures 5.4 and 5.5. The Ca^{2+} complex showed a regular pentagonal bipyramid geometry with bond lengths of 2.4 Å for Ca–O which matched with the structure obtained from *abinitio* molecular orbital calculations (Katz et al., 1996). The Mg^{2+} complex showed an octagonal geometry for the first solvation shell, consisting of 6 water molecules and a bond length of 2.1 Å for Mg–O. The second solvation shell consisted of 12 water molecules hydrogen bonded to the inner–shell water molecules. The geometry of this complex could be represented as a pentagonal dodecahedron arrangement of water molecules enclosing the $[Mg(H_2O)_6]^{2+}$ octahedron and it matched very well with results obtained from molecular orbital theory and DFT (Markham et al., 2002).

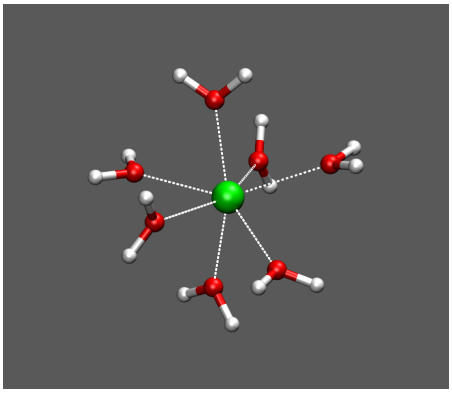


Figure 5.4: Quantum optimized structure of the Ca^{2+}–water complex. The geometry represents a regular pentagonal bipyramid arrangement of the first shell water molecules.

The single point energy is the energy of the molecule at a specific geometry which in our case was the optimized structure and it was evaluated by sequential addition of water molecules to the solvation shell:

$$E_{tot} = E(M^{2+}) + nE(H_2O) - E\{M(H_2O)_n^{2+}\} \tag{5.4}$$

$$\Delta E = E\{M(H_2O)_{n-1}^{2+}\} + E(H_2O) - E\{M(H_2O)_n^{2+}\} \tag{5.5}$$

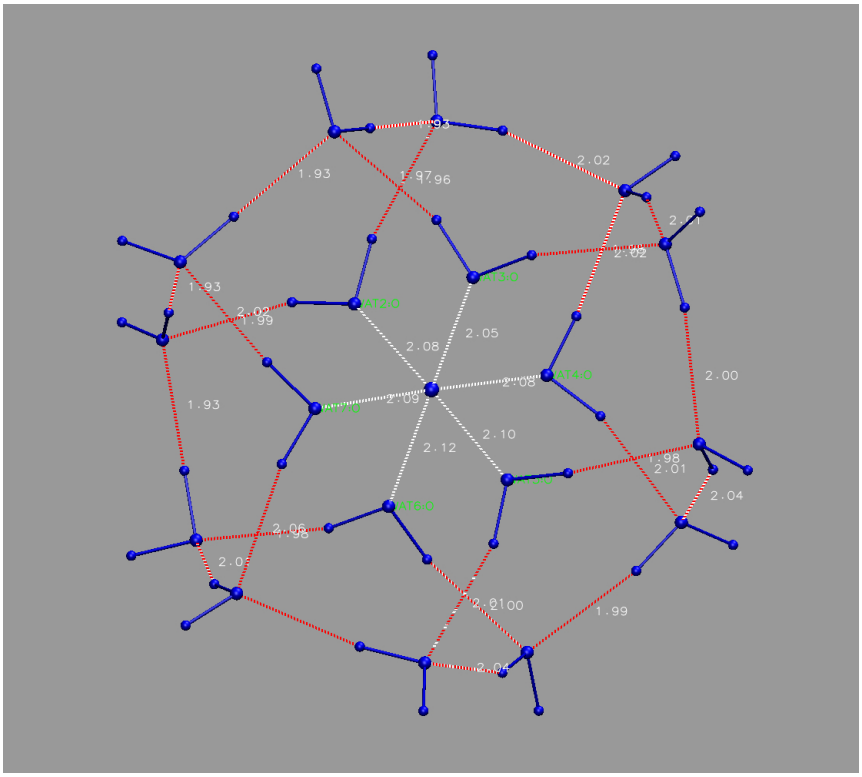


Figure 5.5: Quantum optimized structure of the Mg^{2+}–water complex. The geometry represents a pentagonal dodecahedron second shell enclosing an octahedral first shell.

(Pavlov et al., 1998) performed DFT calculations on these complexes and reported the single point energy values of -234.4 kcal/mol for ($[Ca(H_2O)_6]^{2+}$) and -460.8 kcal/mol for $[Mg(H_2O)_6]^{2+}(H_2O)_{12}$, which we used to compare to our calculated value from molecular mechanics. The energy values for each complex were observed for different charge distributions on the multisite cation model and they were matched to the aforementioned values.

3. Fitting the Hydration Free Energy

The hydration free energy of a cation was defined as the energy required in transferring a cation from gas phase to an aqueous solution at a standard state: $Ca^{2+}(g) \rightarrow Ca^{2+}(aq)$. This gave the change in energy as: $\triangle G = G_{aq} - G_g$. This energy was then broken down in the following steps: discharging the gaseous ion, hydrating the

resulting neutral species and then recharging it in solvent to get: $\triangle G = G_{disch} + G_{neutr} + G_{charge}$.

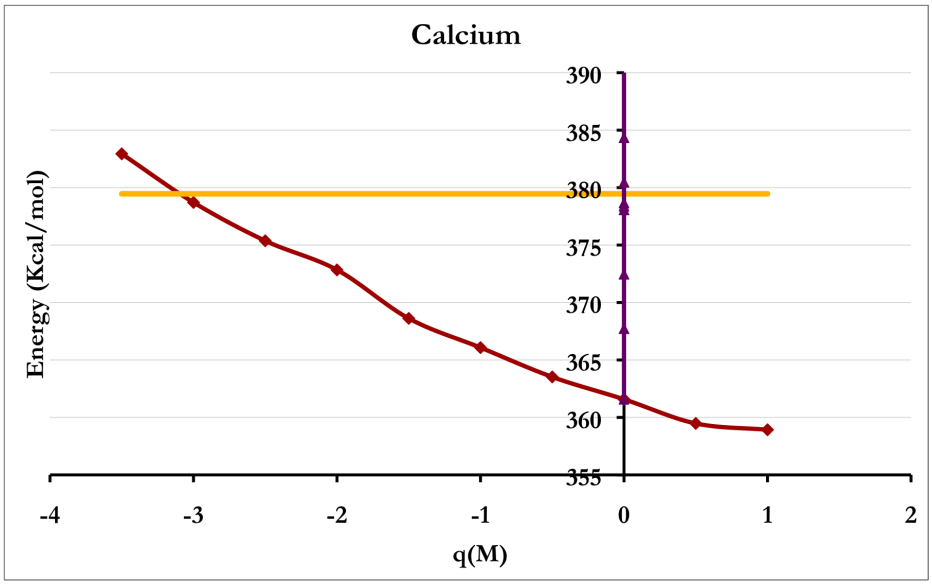


Figure 5.6: Charge on the central atom M plotted against the solvation energy obtained from the thermodynamic integration calculations of the Ca^{2+} cation. The yellow line is the experimental solvation energy (ΔG = -379.46 kcal/mol (Noyes, 1962)). The red line represents the energy values for variation in charge distribution. The purple line along the vertical axis at $q(M) = 0$ shows the energy values obtained when varying the Lennard–Jones parameters. The Energy values on the y-axis are negative (< 0).

Thermodynamic integration routine of AMBER8 was used to calculate the energy of each transition mentioned above. The method calculates the transition from the intial state to the final state by changing the coupling parameter λ from 0 to 1. The total change in energy ΔG is obtained by numerical integration of $dU/d\lambda$ over λ values (Details of this method are given in Chapter 2). For our calculation, the cation was first solvated in a truncated octahedral periodic boundary cell with a 24 Å TIP3P water padding. The large water padding was used to get rid of the artifacts of using PME in ion-water simulations (Piquemal et al., 2006). No counter ions were used and the long range forces were calculated using Particle Mesh Ewald (PME). Covalent bonds to hydrogens were constrained with SHAKE and the time step for MD integration was 1 fs. By not using any counter ions in this system, a net positive charge was carried in a periodic system which seems unphysical and

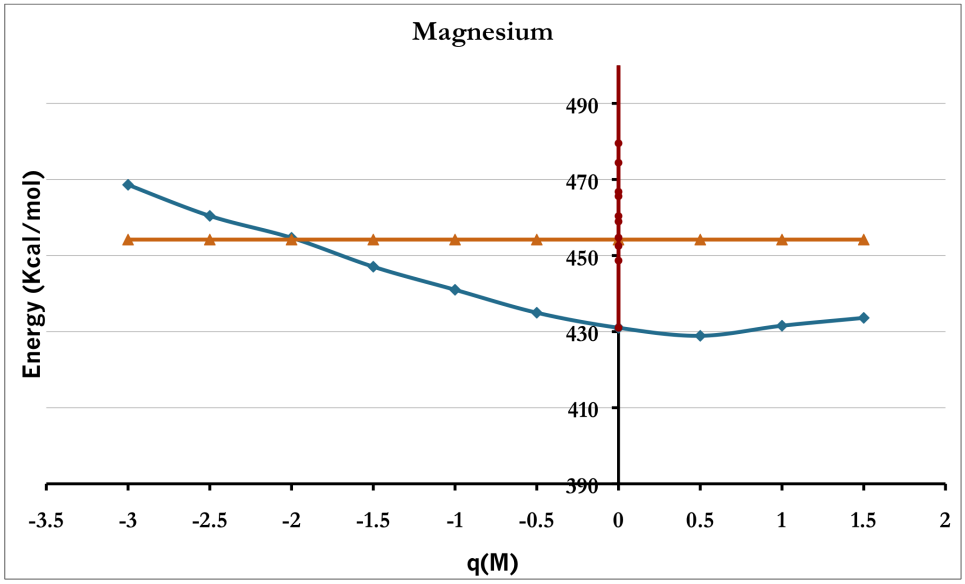


Figure 5.7: Charge on the central atom M plotted against the solvation energy obtained from the thermodynamic integration calculations of the Mg^{2+} cation. The orange line is the experimental solvation energy (ΔG = -454.2 kcal/mol (Noyes, 1962)). The blue line represents the energy values for variation in charge distribution. The magenta line along the vertical axis at $q(M) = 0$ shows the energy values obtained upon varying the Lennard–Jones parameters. The Energy values on the y-axis are negative (< 0).

unrealistic. However, researchers have shown earlier that the calculation of ionic hydration which involves changing the net charge of the system give accurate values for the aforementioned conditions (Hummer et al., 1996; Darden et al., 1998) so we followed the same procedure.

The NPT ensemble was minimized and equilibrated to 298 K, for 100 ps to get rid of any steric clashes. After this the simulations were continued for production run at every λ values. The octahedral cell volume for the system was about 113749 $Å^3$ and it contained 3993 water molecules. The time for every λ run was optimized at a value where $dU/d\lambda$ became stable. This time was chosen as the length of all production runs (200 ps). Since the cation was already solvated in water, the calculation of hydration energy involved going in the reverse direction. First the cation was discharged in water and then disappeared to find the entire change in energy. The values of λ were: 0.00922, 0.04794, 0.11505, 0.20634, 0.31608, 0.43738, 0.56262, 0.68392, 0.79366, 0.88495, 0.95206, 0.99078. The $dU/d\lambda$ obtained for the change in λ was fit with a

polynomial and then integrated to find the total change in energy ΔG. The ΔG values were compared with the corresponding experimental values, -379.46 kcal/mol for Ca^{2+} and -454.2 kcal/mol for Mg^{2+} (Noyes, 1962). Figure 5.6 and Figure 5.7 show the energy values obtained for different charge distributions and different Lennard–Jones parameters of the multisite cation models.

4. Selection of the Parameters

The values for $q(M)$, $q(D)$, A_M, B_M, A_D, B_D were found such that the experimental energies were accurately reproduced. The calculations shown above gave multiple values for these variables which satisfied either all or some of the criteria. The values which minimized the difference between the calculated and experimental energy for all the three measures were picked. The selected values have been tabulated in Table 5.2.

Cation	$q(M)$	$q(D)$	A_M	B_M	A_D	B_D
			((kcal/mol)Å^{12})	((kcal/mol)Å^{6})	((kcal/mol)Å^{12})	((kcal/mol)Å^{6})
Ca^{2+}	0.00	0.28	233.2	36.0	0.05	0
Mg^{2+}	0.00	0.33	27	16.7	0.05	0

Table 5.2: Non–bonded parameters of the multisite cations after refinement

5.5 Cation–Water Simulations

This section presents results of the simulation of the multisite cation model in water. The parameters obtained above were used in these simulations and the results were were analyzed for structural and dynamical properties of the cation environment to compare with the corresponding experimental values

5.5.1 Simulation Protocol

Each multisite cation model, Ca^{2+} and Mg^{2+} was used to run a simulation in water. For each simulation the cation was first solvated in a truncated octahedral periodic boundary cell with 24 Å TIP3P water padding. A 10 Å direct-space cutoff was used for short range Lennard-Jones and electrostatic interactions, long-range electrostatics were treated by Particle Mesh Ewald method (PME) and the covalent bonds to hydrogen atoms were constrained with SHAKE. The system was first minimized by using a hybrid method of switching between steepest descent and conjugate gradient method after every 10 steps and then equilibrated to 298 K, heating gradually in 50K/10ps steps for 200 ps. These were followed by the production phase of the system in an NPT ensemble for 50 ns. All calculations used 1 fs time step in the equilibration phases and 2 fs time step in the production phase. The temperature control in all simulations was performed by Langevin dynamics with a collision frequency of 3 ps^{-1}. In a Langevin dynamics temperature control the kinetic energy is adjusted to be correct for the harmonic oscillator. The physical collision frequencies are in the range of 50 ps^{-1} for liquid water but in simulations it is better to use much smaller values (between 2 to 5 ps^{-1}) for better sampling and stability of the system (Pastor et al., 1988).

5.5.2 Results

The structural and dynamical properties of the cation environment were matched with the experimentally obtained values to measure how accurately the model was able to represent ion-water interactions. Table 5.3 and 5.4 shows the results from the

Ca^{2+}–water and Mg^{2+}–water simulations. The values from the X-ray crystallography, neutron–scattering experiments and also other simulations studies were used to compare. The results from this work showed that they were in good agreement with the experimental values. Also, when compared with recent cation-water simulations in a polarizable force field, AMOEBA (Ponder et al., 2010), it was seen that the multisite cation model was able to reproduce the properties as accurately as the simulations in a polarizable force field (Piquemal et al., 2006). The interesting feature was that the multisite cation model was able to yield those values in a non-polarizable force field and the simulations were not even computationally intensive. Each structural and dynamical property of the ion environment is discussed below:

Cation environment: Ca^{2+}–water simulations				
Properties	Multisite Cation Model	Experimental	AMOEBA simulation	Other simulation
Coordination number	7.0–8.0	6.0–8.0	7.2–7.7	6.2–7,8-9
First solvation shell peak (Å)	2.35	2.41–2.44	2.42–2.56	2.43–2.44

Table 5.3: Structural and dynamical properties for Ca^{2+}–water simulations. Experimental values obtained from (Jalilehvand et al., 2001; Badyal et al., 2004; Licheri et al., 1975, 1976; Hewish et al., 1982; Yamaguchi et al., 1989). Other simulation data from (Naor et al., 2003).

Cation environment: Mg^{2+}–water simulations				
Properties	Multisite Cation Model	Experimental	AMOEBA simulation	Other simulation
Coordination number	6.0	6.0	6.0	6.0
First solvation shell peak (Å)	2.07	2.09	2.08	2.13

Table 5.4: Structural and dynamical properties for Mg^{2+}–water simulations. Experimental values from (Neely and Connick, 1970; Tongraar and Rode, 2005) and other simulation values obtained from (Lightstone et al., 2001).

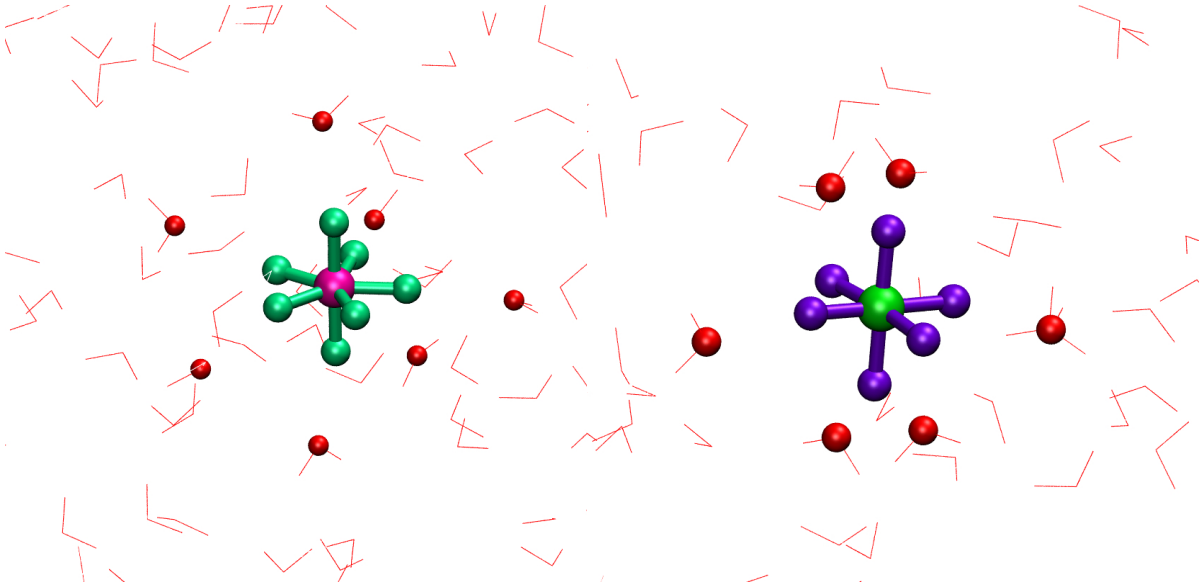


Figure 5.8: Ca^{2+} ion in water

Figure 5.9: Mg^{2+} ion in water

Coordination Number

Calcium–water simulation showed a coordination number of 7 most of the time (70%) (Figure 5.8). The remaining snapshots (30%) had a coordination number of 8. Both these values matched well with the experimental data (Licheri et al., 1975, 1976; Hewish et al., 1982; Yamaguchi et al., 1989) which have shown that the coordination number of Ca^{2+} fluctuates between 6 to 8. We would like to point out here that the geometry of the 7 dummy centers in the Ca^{2+} multisite model did not force the coordination to 7. The model is flexible enough to allow realistic geometries.

The Mg^{2+}–water simulations showed a coordination number of 6 (Figure 5.9) for all snapshots which is what was observed experimentally as well. X-ray crystallography data has shown that Mg^{2+} ion forces a strict coordination of 6 members in its first solvation shell which was maintained in our simulations.

Radial Distribution Function (RDF) of the Cation

The radial distribution function describes the structural and dynamical properties of the water molecules around the Ca^{2+} ion. In our simulations (Figure 5.10) the first

sharp peak occurred at 2.4 Å. This position agrees with the average distance Ca–O distance in water (2.4 Å) deduced from X-ray and neutron scattering experiments (Licheri et al., 1975, 1976; Hewish et al., 1982; Yamaguchi et al., 1989; Fulton et al., 2003; Gaspar et al., 2004; Megyes et al., 2004).

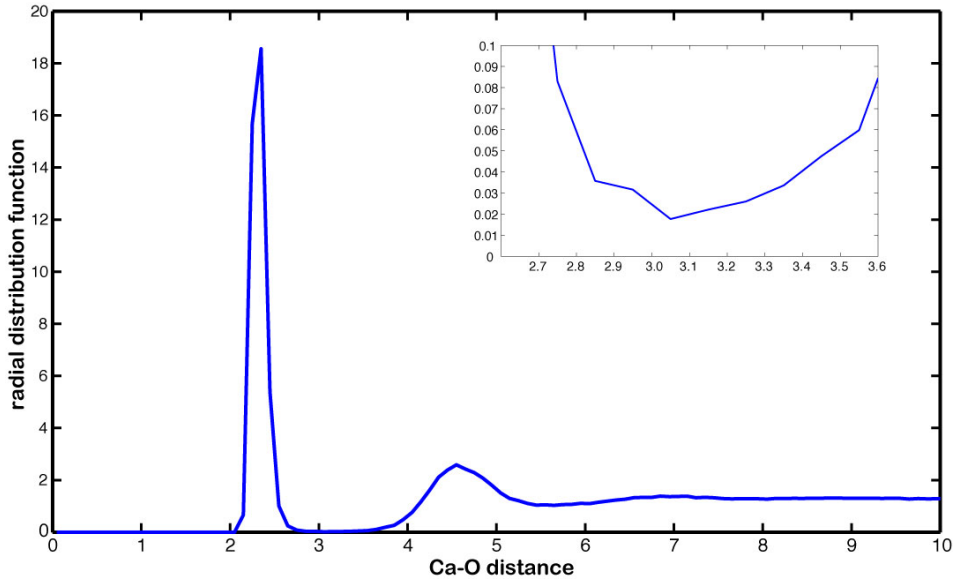


Figure 5.10: Radial distribution function of the Ca^{2+}–water simulation

The interesting observation was the occurrence of the second peak. Earlier simulations performed using the spherical ion found it challenging to witness the second peak in the RDF of Ca^{2+} ions. The non–appearance of this peak indicated that those models were not sufficient to determine the orientation of the second shell waters. Simulations in the polarizable force field and a specialized potential were able to observe the second shell (Piquemal et al., 2006). The results shown here matched very well with these findings indicating that by using a multisite cation model, the geometry and orientation of both the first shell and the second shell atoms can be determined accurately.

For Mg^{2+} the radial distribution function showed a peak at 2.07 Å (Figure 5.11) which matched well to 2.08 Å value seen experimentally (Tongraar and Rode, 2005). The second peak is observed at 4.35 Å, which matches with the value of 4.2 Å observed in the quantum studies (Markham et al., 2002).

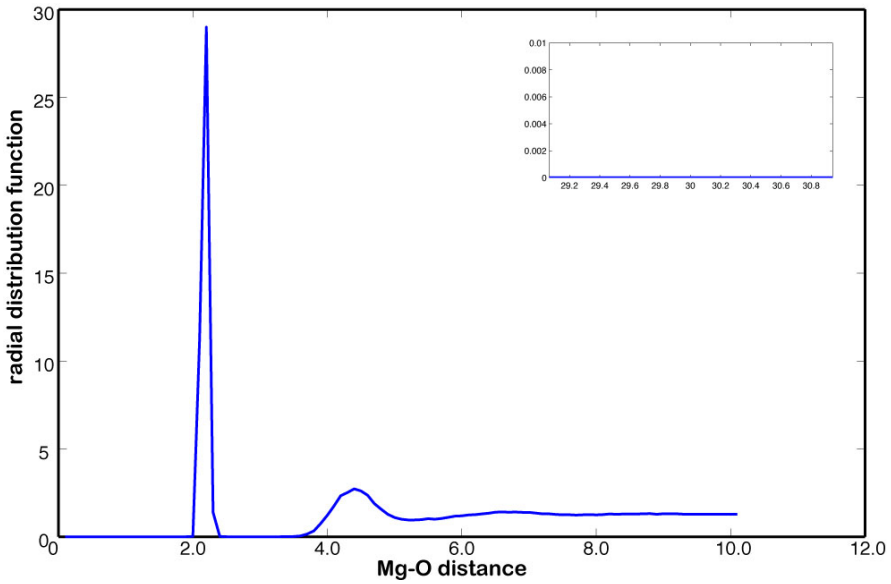


Figure 5.11: Radial distribution function of the Mg^{2+}–water simulation

Water Exchange Rates

Earlier simulations had found it challenging to observe high exchange rates from the first solvation shell of Ca^{2+} due to stability issues (Owczarek et al., 2007). The simulations with the multisite cation model was able to see significant exchange of water molecule. This can be observed in the RDF profile in Figure 5.10. The inset shows that radial distribution value of water molecules at the first dip never goes to zero, revealing that there is an exchange between the first and second solvation shells. These results also agreed well with the recent simulation in the polarizable force field (Piquemal et al., 2006). The higher exchange are required to allow the coordination number to shuttle between 7 and 8 which has been experimentally observed for the Ca^{2+} (Badyal et al., 2004). For Mg^{2+} -water simulation, no water exchange was observed (Figure 5.11). In experiments the Mg^{2+} ion has shown very long exchange rates, in the order or seconds which is very difficult to observe in a 50 ns simulations (Neely and Connick, 1970).

5.6 Simulations of Proteins Bound with Cations

The purpose this section is to test the multisite cation model in simulations with ion binding proteins. The prototypical ion binding proteins which were picked for these simulations were: Ca^{2+} binding protein, Calbindin and Mg^{2+} binding protein, CheY. These proteins were used to run simulations with the Ca^{2+} and Mg^{2+} multisite cations and the structural properties of the binding sites were studied for accuracy.

5.6.1 Calcium Bound Protein – Calbindin

Calbindin belongs to the EF-hand class of proteins which is the most common family of Ca^{2+} binding proteins. It is the smallest member of this family, containing only 75 residues and possessing a pair of functional EF hand loops, both of which bind to Ca^{2+} with a high affinity. The association constants lie between 2 x 10^6 and 6 x 10^8 M^{-1}. Calbindin represents a suitable system to work with as it has been extensively characterized both experimentally and structurally. The crystal structure of the protein with Ca^{2+} bound to both the sites has been determined by X-ray crystallography (Szebenyi and Moffat, 1986). In addition, solution structures have been determined for doubly loaded, singly loaded state and apo state of the protein (Akke et al., 1991, 1995). The three dimensional structure of the protein can be seen in Figure 5.12.

Previous computational studies on this protein suffered several artifacts. The doubly loaded crystal structure of this protein was used to run simulations using the extended atom and all atom representations (Ahlstroem et al., 1989; Koerdel and Teleman, 1992). In both these simulations, periodic boundary conditions were used with a cutoff of 10 Å for non-bonded interactions and it was found that the local coordination patterns of the Ca^{2+} ion and the secondary structure of the protein were significantly altered. In particular, there was a wrong coordination of Glu60 with the Ca^{2+} ion of Site I. These artifacts could be corrected by using a spherical solvent boundary potential (SSBP) with an extended electrostatics procedure and no cutoff in some later simulations (Marchand and Roux, 1998). This method was more accurate but was limited to small systems and computationally more expensive. In this section,

it was found that by using a Ca^{2+} multisite model in the simulations the previously faced artifacts were removed. Additionally, the simulations could be performed with the default periodic boundary conditions and with systems of any size.

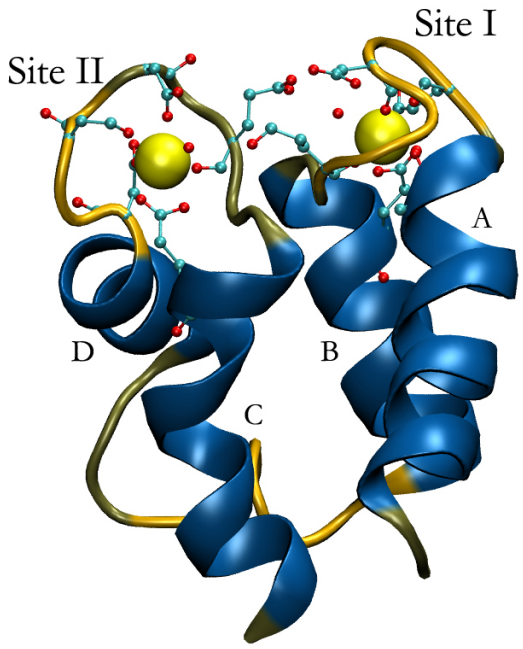


Figure 5.12: Calbindin crystal structure

5.6.2 Magnesium Bound Protein – CheY

CheY is the *E. coli* chemotaxis signal transduction protein that is activated through phosphorylation (Wadhams and Armitage, 2004). Depending upon the state of phosphorylation, the binding affinity of CheY to FliM is modulated which in turn determines the rotational pattern of the flagella (clockwise vs counter-clock-wise) (Toker and Macnab, 1997). The direction of the flagellar motors eventually regulate the swimming behavior of the bacteria from being tumbled or smooth. The crystal structure of the Mg^{2+} bound CheY has been determined and is shown in Figure 5.13 (Stock et al., 1993). The Mg^{2+} binding site in this protein is formed by Asp13, Asp57 and Asp59 which are present towards the periphery. These residues form one hemisphere of the coordination sphere. The other hemisphere is filled by three oxygen atoms coming from the water molecules.

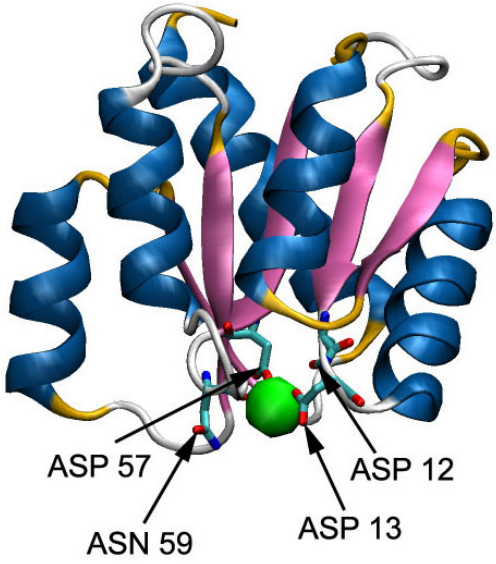


Figure 5.13: CheY with Mg^{2+} multisite cation

5.6.3 Simulation Protocol

Two simulations were run for each protein, one using a spherical ion and the other using the multisite cation model with identical simulation conditions. The starting structure was first solvated in a truncated octahedral periodic boundary cell with 12Å TIP3P water padding. Na^+ and Cl^- counterions were used to first neutralize the system and then to bring the salt concentration of the solution to 50 mM. A 10 Å direct-space cutoff was used for short range Lennard-Jones and electrostatic interactions. Long-range electrostatics were treated by Particle Mesh Ewald method (PME) and the covalent bonds to hydrogen atoms were constrained with SHAKE. The system was first minimized by using a hybrid method of switching between steepest descent and conjugate gradient method after every 10 steps and then equilibrated to 298 K, heating gradually in 50K/10ps steps for 200 ps. The equilibration step was divided into two steps: the initial step involved 10 kcal/mol restraints on all protein atoms to keep them from moving away from their starting positions and letting the water box equilibrate for 200 ps. The second step involved NVE simulation without any restraints on any atom for 200 ps. These were followed by the production phase of the system in an NPT ensemble for 80 ns. 1 fs time step was used for the equilibration phases and 2 fs for the production run. The temperature control in all simulations was performed by Langevin dynamics with a collision frequency of 3 ps^{-1}.

5.6.4 Results

Structure of the Binding Sites in Calbindin

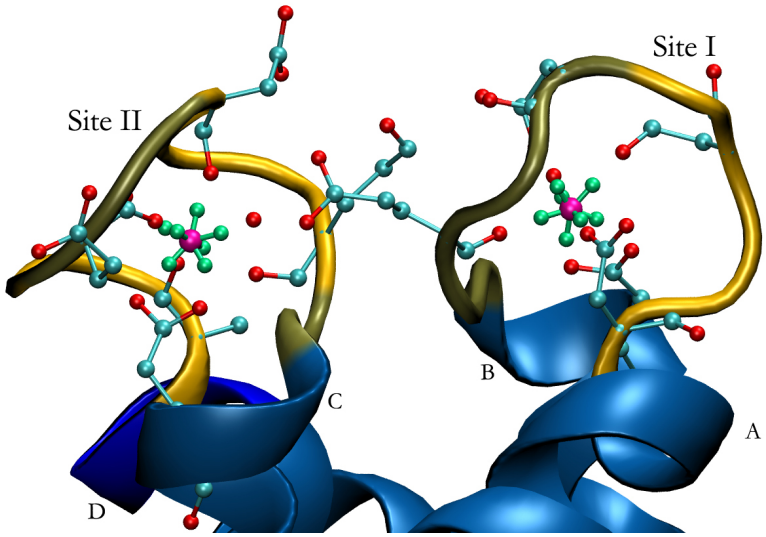


Figure 5.14: Calbindin with mutisite Ca^{2+} ion

The crystal structure of Calbindin had two binding site: Site 1 (pseudo EF-hand, 14 residues) from Ala14 to Glu27 and Site II (standard EF-hand, 12 resdiues) from Ala54 to Glu65. In an EF-hand protein, the Ca^{2+} is known to be coordinated by seven oxygen ligands coming from the 12 residue loop (Strynadka and James, 1989). These coordinations are provided by five side chain carboxyl oxygens, one backbone carbonyl oxygen and one water oxygen. The Site II of Calbindin follows this scheme with the 12 residues from Ala54 to Glu65 enclosing the Ca^{2+} ion, and is thus a standard EF-hand motif. Site I, however, is formed by a 14 residue loop, from Ala14 to Glu27, and is therefore termed as the pseudo EF-hand motif. The difference mostly occurs in the coordinating pattern with the oxygen atoms now being contributed by four backbones, two side-chains and one water molecule.

The structure of the binding site in our simulations (Figure 5.14) showed that during the entire simulation the cation stayed stably bound to both the binding site. The comparison of the average coordination patterns of the both sites with the experimental values (Svensson et al., 1992) can be found in Table 5.5. The distances

Cation-oxygen distances (Å)		
Residues	Experimental	Multsite cation MD
Site I		
Ala14 O	2.33	2.31 ± 0.07
Glu17 O	2.47	2.33 ± 0.08
Asp19 O	2.25	2.32 ± 0.08
Gln22 O	2.36	2.38 ± 0.10
Glu27 OE1	2.42	2.29 ± 0.07
Glu27 OE2	2.60	2.23 ± 0.05
Water OW	2.43	2.33 ± 0.07
Site II		
Asp54 OD1	2.07	2.20 ± 0.04
Asp54 OD2	–	4.16 ± 0.20
Asn56 OD1	2.15	2.26 ± 0.06
Asp58 OD2	2.11	2.22 ± 0.05
Glu60 O	2.19	2.26 ± 0.06
Glu65 OE1	2.15	2.28 ± 0.07
Glu65 OE2	3.30	2.25 ± 0.06
Water OW	2.53	2.31 ± 0.06

Table 5.5: Coordination distances for Calbindin binding sites. Experimental values obtained from (Svensson et al., 1992).

matched very well with the experimental values and the pattern also matched with those observed by (Marchand and Roux, 1998). Ion distances from Asp19 and Gln22 showed a slight increase while the Glu27 side chain was seen to have gone closer to the ion. An extra water molecules was seen to enter the coordination sphere of Site I which was also observed in the previous simulation (Marchand and Roux, 1998). This water molecule, however, did not stay bound for a long time and it left the site mid-way during the simulation. This indicated that the ion did not really require this water molecular to complete its coordination.

The interesting observation here was for the side chain of Glu60. This side chain was seen to coordinate with the Ca^{2+} ion of Site I in all earlier simulations (Ahlstroem et al., 1989; Koerdel and Teleman, 1992) and it was corrected when the simulations involved SSBP and extended electrostatics potential (Marchand and Roux, 1998). In this work it was observed that the Glu60 did not wrongly coordinated with both ions, matching the experimental observations.

Site II residues efficiently maintained the EF-hand binding site conformation and showed coordination patterns which matched very well with the experiments. There was a slight increase in all distances except for Glu65 side chain which was pulled in, to coordinate strongly. All the previous simulations saw inclusion of an extra ligand, Asp54 OD2 in this Site II which was believed to be caused by neglect of induced polarization terms in the potential function. Here, the simulations maintained the experimental structure and did not see this extra coordination indicating that the multisite model was successful in maintaining the coordination geometry.

Structure of the Binding Site in CheY

The binding site in the Mg^{2+} binding protein CheY consists of six coordinating oxygen atoms, three from the protein: Asp13, Asp57 and Asp59 and the remaining three from water molecules. The coordination distances for the binding site obtained from the multisite cation simulations are reported in Table 5.6. In general, the simulations saw an increase in coordination towards the protein with distances becoming less than the experimental values. The ion remained bound in a stable fashion throughout the length of the simulation.

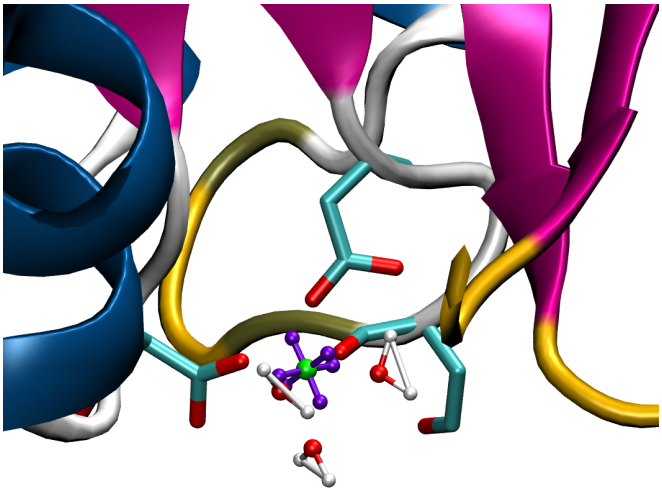


Figure 5.15: Binding site in CheY

Cation-oxygen distances (Å)		
Residues	Experimental	Multsite cation MD
Asp 13 OD1	1.99	1.88 ± 0.05
Asp 57 OD2	2.09	1.88 ± 0.05
Asn 59 O	2.09	2.02 ± 0.08
Water 1 OW	2.15	1.99 ± 0.07
Water 2 OW	1.92	1.98 ± 0.07
Water 3 OW	1.92	1.99 ± 0.07

Table 5.6: Coordination distances for CheY binding site. Experimental values obtained from (Stock et al., 1993).

5.7 Selectivity between Calcium and Magnesium ions in a Calcium Binding Protein

The purpose of this section is to show that the multisite cation model can be used to reproduce accurate thermodynamical values for complex mechanisms such as identifying selectivity between similar ions. Ca^{2+} and Mg^{2+} are the most relevant cations, physiologically. While Mg^{2+} forms an important cofactor of enzymes catalyzing reactions underlying transcription, translation, and replication of nucleic acids, Ca^{2+} is involved in various signaling pathways and functions such as muscle contractions, neurotransmitter release, etc. Both these ions are small and closed-shell metal ions with the same net charge. Despite their similarities, these ions have distinct functions to cater to. Mg^{2+} is present in mM concentration and is used for a wide range of purposes. Ca^{2+}, on the other hand, is present in micromolar range but is detected by specialized proteins which respond to sudden rise in its concentration and spikes. The binding affinities for Ca^{2+} in these cases are seen to be 10^2 to 10^4 fold higher than for Mg^{2+}. Structural studies have shown that the Ca^{2+} binding proteins are able to achieve this selection by exploiting the subtle differences between the two cations. The two ions differentiate in the ionic radius and the shape of their coordination polyhedra. Magnesium has a smaller ion radius than Ca^{2+} and it also favors strict sixfold, octahedral coordination geometry (Falke et al., 1994). Calcium on the other hand is flexible and can have from 6 to 8 coordinates. In most of the cases it has been seen to attain a pentagonal bipyramid geometry, coordinating with 7 atoms (Strynadka and James, 1989).

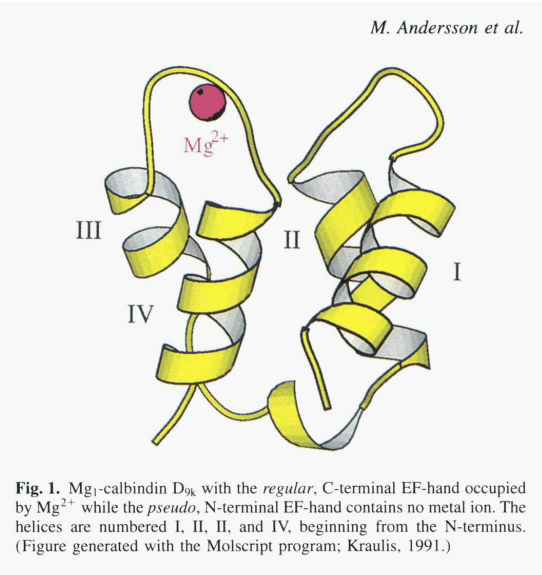


Fig. 1. Mg_1-calbindin D_{9k} with the *regular*, C-terminal EF-hand occupied by Mg^{2+} while the *pseudo*, N-terminal EF-hand contains no metal ion. The helices are numbered I, II, II, and IV, beginning from the N-terminus. (Figure generated with the Molscript program; Kraulis, 1991.)

Figure 5.16: Mg^{2+}–bound crystal structure of Calbindin. Reprinted by permission from John Wiley and Sons (Andersson et al., 1997).

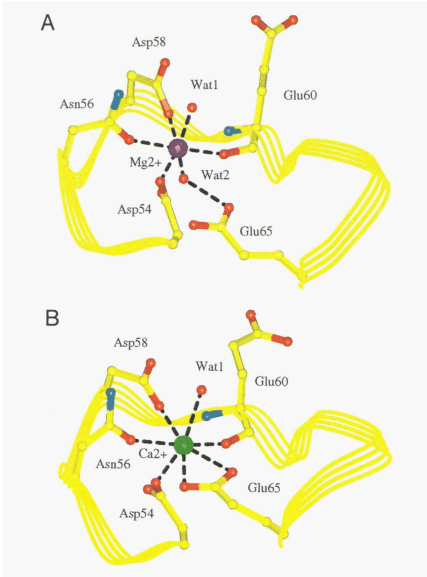


Figure 5.17: Ca^{2+} and Mg^{2+} bound crystal structures of Calbindin Reprinted by permission from John Wiley and Sons (Andersson et al., 1997).

One of these specialized class of proteins, which selectively bind to Ca^{2+} ions are EF-hand proteins. There are some in this class which are flexible and can bind to either ion, but most of them are principally selective to Ca^{2+}. EF-hand proteins are involved in a wide variety of physiological processes, including signaling, cell cycle regulation, muscle contraction, vision, etc. The proteins belonging to this family contain homologous Ca^{2+}-binding sites within a characteristic helix-loop-helix motif called the EF-hand (da Silva and Reinach, 1991). The EF hand, as traditionally defined, is a loop containing 12 sequential residues, starting with the first coordinating residue and ending at the last coordinating residue and a total of 6 residues participating in the coordination. A pair of such loops is generally found back to back arranged in an anti-parallel manner. Members of the EF-hand family include Calmodulin, Calbindin, Parvalbumin, Troponin C, Myosin regulatory chains and more. Several of these members, such as Calmodulin, undergo a large conformational change upon binding to Ca^{2+} ions. Other members such as Calbindin and Parvalbumin do not demonstrate such effect.

Calbindin is a small globular protein containing two EF-hand loops. The protein participates in Ca^{2+} transport and is known to bind to two Ca^{2+} ions with positive

cooperativity. Studies have suggested that this protein also participates in Mg^{2+} absorption in the intestine (Hemmingsen et al., 1994). (Andersson et al., 1997) in 1997 looked at Mg^{2+} binding in this protein and found that Mg^{2+} could bind to the protein but with much lower affinities than Ca^{2+}, and that too only to Site II (Figure 5.16 and 5.17). The thermodynamic and structural explanation to such selectivity is still unknown. Here, the aforementioned phenomenon has been investigated through molecular dynamics simulations.

MD simulations can be successfully employed for such a study as they allow convenient visualization of the intermediates of the mechanism and can also allow a quantitative analysis. The simulations, however, have several limitations in the way the ion-protein interactions are calculated and hence sometimes become challenging. Using a spherical representation of the cation in such a study makes it very difficult for the protein to identify the subtle differences. Additionally, it's still questionable which parameters of the ions are good for a particular system. (Allouche et al., 1999) used spherical ion representation and found accurate relative binding energies of Ca^{2+} vs Mg^{2+} binding in Parvalbumin, but they had to put constraints between the protein and the cations. (Cates et al., 2002) did a similar study where they used improved parameters for the cations such that they did not need external constraints. However, they still required artificial tethering of the coordinating water molecule and reparamterization of a coordinating Ser, to get the correct behavior. The multisite cation model can prove advantageous in such a study as it inherently captures the subtle differences between these cations by defining their respective coordinating geometries. Thus, in this section the multisite cation representation has been used with the protein Calbindin to investigate the selectivity of Ca^{2+} over Mg^{2+}. The results showed that by using our method the relative binding energies of Ca^{2+} and Mg^{2+} matched accurately with the experimentally obtained values. Moreover, the simulations were also able to show the structural transition of the Ca^{2+} coordinate geometry (pentagonal bipyramid) to the Mg^{2+} coordinate geometry (octagonal) and vice versa.

5.7.1 Methods

Calculating relative binding free Energies

The method involved finding the relative binding energy $\Delta\Delta G$, for the processes:

$$PROT + Ca \rightarrow PROT \cdot Ca(\Delta G_1) \tag{5.6}$$

$$PROT + Mg \rightarrow PROT \cdot Mg(\Delta G_2) \tag{5.7}$$

Where $\Delta\Delta G = \Delta G_2 - \Delta G_1$. The thermodynamic cycle perturbation method, proposed by (Lybrand et al., 1986) was used to obtain $\Delta\Delta G$. Figure 5.18 shows the closed thermodynamic cycle where $\Delta\Delta G = \Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3$. The thermodynamic integration method in AMBER8 was used to calculate ΔG_4 and ΔG_3 (Section 2.4).

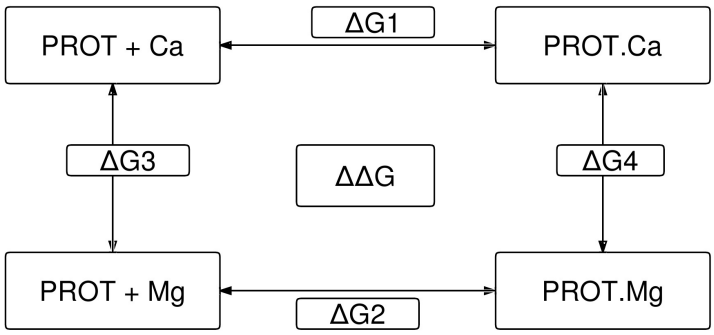


Figure 5.18: Thermodynamic coupling cycle for finding relative binding energy of Ca^{2+} to Mg^{2+} binding to a protein

Initial Structures

The starting coordinates were the Ca^{2+} bound form of Calbindin originated from the 2.3 Å atomic resolution crystal structure, 3ICB (Moffat and Moffat, 1986). The Ca^{2+} ion in the crystal structure was replaced with the multisite model. For the reverse run the starting coordinates were of the Mg^{2+} bound form of Calbindin, a 1.6 Å atomic

resolution crystal structure, 1IG5 (Andersson et al., 1997). Here, the Mg^{2+} ion was replaced by the corresponding multisite model. Additionally, parallel simulations for both transitions were run using the spherical ion representation. AMBER/parm94 force field parameters were used for the protein and for the multsite cations the parameters obtained earlier in this chapter were used. The parameters for the spherical ion were obtained from Aqvist's paper (Aaqvist, 1990).

Simulation Protocol

Two sets of thermodynamic integration calculations were setup for each step to find the energy change: first using the multisite cation and the other using the spherical ion. Identical simulation parameters were used for both the simulations. Additionally, reverse simulations were set up for each case and with both types of ion representations. The spherical ion simulations were setup to compare the change in energy obtained from the multisite cation model. Each set of simulation was done three times and the mean energy change from these was recorded as the final value. The values obtained from these calculations were compared with the experimentally obtained values in (Andersson et al., 1997). AMBER software was used to perform the thermodynamic integration calculations.

For the calculation, the starting structure was first solvated in a truncated octahedral periodic boundary cell with a 12 Å water padding. The water model used here was TIP3P. Counterions were used to neutralize the system such that the results could be compared with the energies corresponding to a salt concentration of < 2 mM. The NPT ensemble was minimized and equilibrated at 298 K, for 200 ps to get rid of any steric clashes. After this the simulations were continued for production run for several λ values ranging from 0 to 1. Every λ step had an additional 200 ps equilibration step before the change in energy was recorded. Long-range electrostatics interactions were treated by Particle Mesh Ewald method (PME). Covalent bonds to hydrogens were constrained with SHAKE and the time step for MD integration was 1 fs. For proteins the octahedral cell volume for the system was about 221488 $Å^3$ and it contained 6078 water molecules, 6 Na^+ ions and 1 Cl^{-1} ions. For every λ value, the change in $dU/d\lambda$ was monitored and a time was optimized at which $dU/d\lambda$ became stable. This time was chosen as the length of all production runs (200 ps).

5.7.2 Results

Calcium to Magnesium Transition in Water

The transition of a Ca^{2+} ion in the solvent to a Mg^{2+} ion was observed with the first solvation shell of the ion changing from seven membered to six membered. The coordination geometry of this shell saw a change from a regular pentagonal bipyramid to an octahedron as seen in Figure 5.19. The change in energy obtained from the thermodynamic integration calculation (-74.70± 0.20 kcal/mol) for this transition was in excellent agreement with the difference in the experimental solvation energies of the two ions (379.46 – 454.2 = -74.70 kcal/mol (Noyes, 1962)).

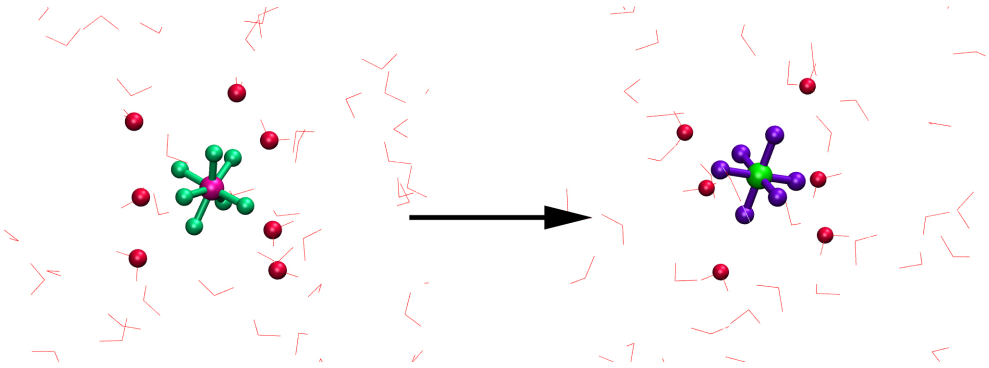


Figure 5.19: Transition of Ca^{2+} to Mg^{2+} in water. The Ca^{2+} ion (green) has seven members in its first solvation shell which becomes six when the ion changes to Mg^{2+} (purple).

Calcium to Magnesium Transition in Protein

The transition from the Ca^{2+} bound protein to a Mg^{2+} bound protein was stable. The starting structure with the multisite cation model exhibited the same coordination with the EF hand as was seen in the crystal structure, a pentagonal bipyramid.

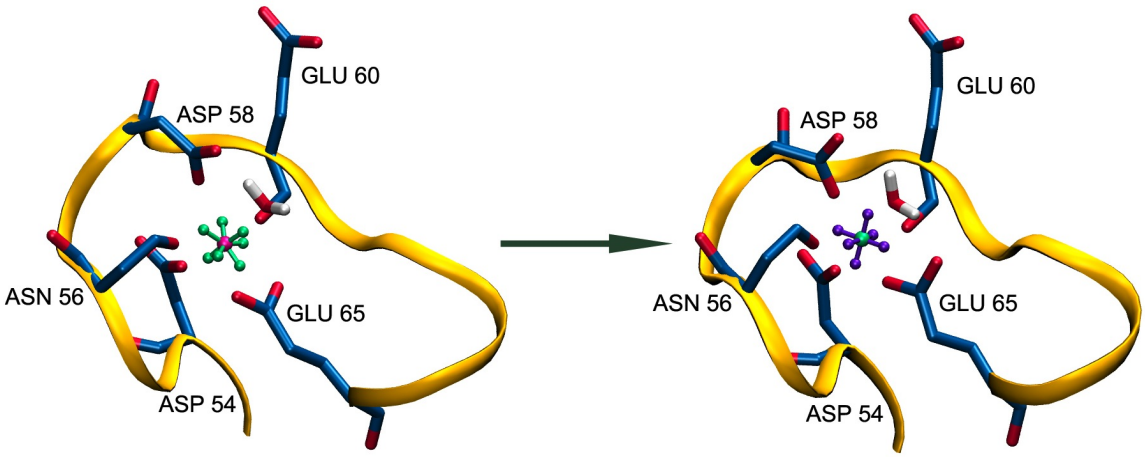


Figure 5.20: Structural change in protein for Ca^{2+} to Mg^{2+} transition. The Ca^{2+} ion (green) has seven members in its first solvation shell which becomes six when the ion changes to Mg^{2+} (purple).

Upon transition to the Mg^{2+}–bound form, the change in coordination geometry to the octahedral form was observed (Figure 5.20). This change in coordination resulted in Glu65 to become monodentate. The remaining coordinating site did not experience any significant change. Upon comparing the final form with the Mg^{2+} bound crystal structure, it was found that the coordination site matched very well with the crystal structure with Glu65 providing a single oxygen atom for coordination with the cation. The only difference between the two structures was the presence of a water molecule in the crystal structure, which was missing from the simulation obtained structure. The reason this water molecule was not seen in the simulations was that the simulation time was not long enough to allow the inclusion of water molecule in the binding site. The Ca^{2+} crystal structure which was used as the starting structure did not have the water molecule to start with.

The similar calculation using spherical ion representation which was run in parallel was seen to be stable as well. The change in coordination from seven members to six members was observed. However, in addition, the conformation of the protein also changed a lot which was not expected.

Next, the energy values obtained from these calculations were examined. The $dG/d\lambda$ values obtained were numerically integrated over λ to get the change in energy. These calculations were done three times and the final energy is reported in Table 5.7 and

Relative binding energy (kcal/mol)			
System	Multisite	Spherical	Experimental Energy
Water	-74.25 ± 0.19	-74.70 ± 0.20	-74.70
Protein	-78.40 ± 0.89	-74.70 ± 0.20	–
$\Delta\Delta G$	-4.15 ± 1.08	-10.47 ± 1.27	-4.5 to -5.76

Table 5.7: Relative binding energy for Ca^{2+} to Mg^{2+} transition.

Figure 5.21. The energy values obtained from the spherical ion simulation are also tabulated. The values showed that the simulation with the multisite cation was able to accurately match the experimental binding energy while the values obtained from the spherical case were much lower in magnitude.

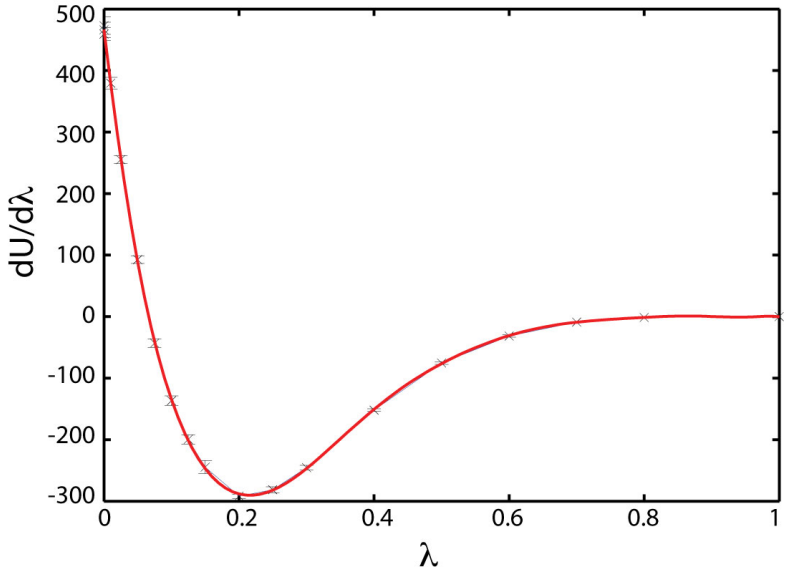


Figure 5.21: λ against dU/dλ values obtained from the thermodynamic integration calculations of Ca^{2+} to Mg^{2+} transition. The integration of this curve gave the total change in energy, ΔG.

Magnesium to Calcium Transition

The simulation for the reverse phenomena, from Mg^{2+} bound protein to the Ca^{2+} bound protein was stable as well. The starting structure of the Mg^{2+} bound form had a water molecule in the coordinating site and it stayed with the ion upon replacing the ion with the multisite cation model and equilibrating the system. The coordination

matched the crystal structure and was octahedral in geometry. Upon transition to the

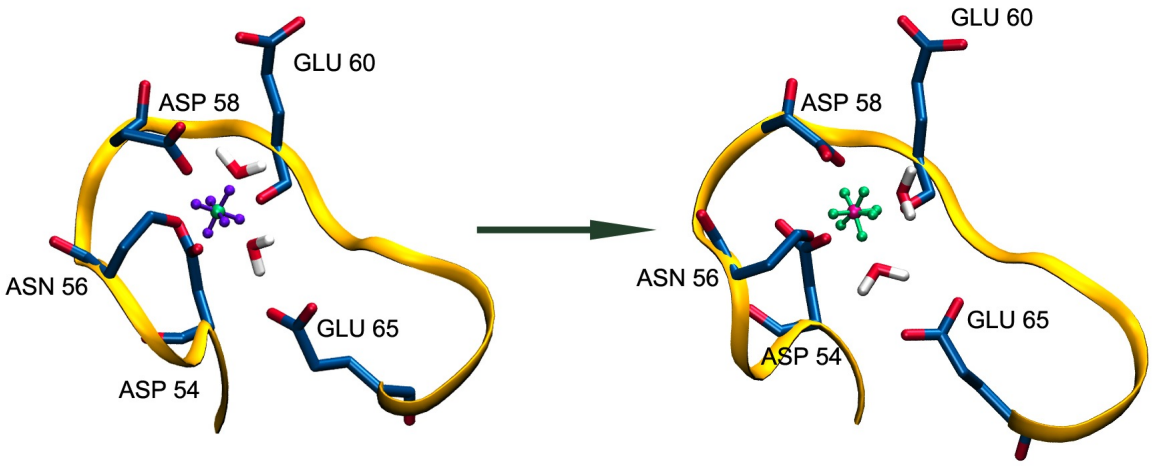


Figure 5.22: Structural change in protein for Mg^{2+} to Ca^{2+} transition. The Mg^{2+} ion (purple) has six members in its first solvation shell which change to seven when the ion becomes Ca^{2+} (green).

Ca^{2+} bound form, the coordination geometry changed to pentagonal bipyramid with Glu65 providing two oxygens for coordination (Figure 5.22). The final form matched with the Ca^{2+} bound crystal structure except for the presence of the water molecule in this case. Again, the reason for this water molecule to have stayed in the binding site is the short length of our simulations. A similar calculation using spherical ion representation was run in parallel. The simulation in this case was also stable and showed a change in coordination from six members to seven.

Relative binding energy (kcal/mol)			
System	Multisite	Spherical	Experimental Energy
Water	74.74 ± 0.14	76.68 ± 0.59	74.70
Protein	79.37 ± 0.78	91.62 ± .04	–
$\Delta\Delta G$	5.128 ± 0.64	14.94 ± 0.26	4.5 - 5.76

Table 5.8: Relative binding energy for Mg^{2+} to Ca^{2+} transition.

The energy values obtained from these calculations are in Table 5.8 and Figure (5.23). The values in the table showed that simulation with the multisite cation was able to accurately match the experimental binding energy. These values also matched the change in energy obtained for the forward case testing the reversibility of the reaction.

The values from the spherical case, however were much lower in magnitude in this case. They did not match either the experimental values or the values obtained from the Ca^{2+} to Mg^{2+} step.

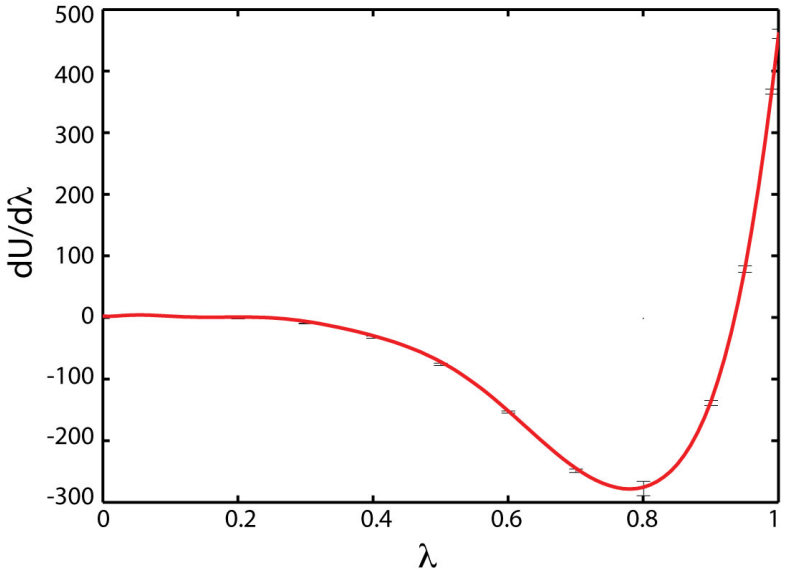


Figure 5.23: λ against dU/dλ values obtained from the thermodynamic integration calculations of Mg^{2+} to Ca^{2+} transition. The integration of this curve gave the total change in energy, ΔG.

5.7.3 Discussion and Conclusions

The results obtained in this chapter show that the multisite cation model improves the ion–protein interactions in molecular dynamics simulations. The simulations of ion-water systems using the multisite cation models are able to capture the accurate structural and dynamical properties of the ion-water interactions, both in the case of Ca^{2+} and Mg^{2+}. The model is seen to accurately represent the geometry of the coordinating atoms and it determines the accurate properties for the second shell atoms as well. The properties obtained are in a very good agreement with the experimental values.

We also compared our values with a recent simulation by (Piquemal et al., 2006) who used the state-of-the-art AMEOBA force field to find the accurate structural

and dynamical parameters of Ca^{2+} and Mg^{2+} in water. It was encouraging to find that the results obtained from using a multisite cation model matched very well with those results, indicating that the use of the multisite model in a non-polarizable force field is able to yield highly accurate results.

The results from the protein simulations showed that by using the multisite cation model in molecular dynamics simulation of ion binding proteins, we were able to maintain the structure of the protein and were also able to accurately represent the binding site structure. In the case of Calbindin, earlier simulation works had shown that the inaccuracies in the Ca^{2+}-binding site occurred due to approximate parameters of Ca^{2+} ion, could be corrected by using a spherical solvent boundary potential (SSBP) with an extended electrostatics procedure and no cutoff (Marchand and Roux, 1998). In this work we showed that we could reproduce these results by using the multsite cation model without the use SSBP and extended electrostatics procedure. In particular, the interaction of Glu60 was corrected and it did not show any wrong coordination with Site I Ca^{2+} ion.

Additionally, our simulations were able to correct another artifact which was seen in all previous simulations. This was the inclusion of Asp54 side chain in the coordination shell of Site I Ca^{2+} ion. All previous simulations, including (Marchand and Roux, 1998), saw an 8 membered coordination at Site I which did not match the experimental results. EF hand motifs are known to be strictly 7 membered (Strynadka and James, 1989) and thus the inclusion of Asp54 side chain in the coordination shell would introduce structural changes in the binding site. We showed that by using the multisite cation model in simulations, Site I maintained the 7 membered coordination and thus maintained the structure of the binding sites. It was believed that the inaccuracies in the earlier simulations occurred due to the neglect of induced polarization in the potential functions, which are known to be significant in the case of ions. As observed in the ion-water simulations, we were able to accurately represent the first and the second coordinating shell and thus were able to account for the polarization effects in the case of proteins as well.

Further advantages of using a multisite cation model could be seen in demonstrating the exceptional property of Ca^{2+}–binding proteins in selecting Ca^{2+} ions over Mg^{2+} ions despite their striking similarities. Because the two ions have almost identical

properties, it has been extremely challenging to show in simulations that proteins had the ability to select one over the other. Researchers had been able to show the change in the coordination pattern, to some extent, but they were never able to reproduce the thermodynamic properties of this selectivity. In our simulations, we showed that by using a multisite cation model, we were able to demonstrate this both structurally and thermodynamically. We used Calbindin as our test system and by using multisite cation model, showed that the EF-hand motif in this site bound to Ca^{2+} ions with a much higher affinity than Mg^{2+} ions. The selectivity was the result of the strict coordination patterns of the 7 membered binding site which could fit the pentagonal bipyramid Ca^{2+} ion very well but could loosely fit the octahedral Mg^{2+} ion, reflecting the difference in their binding affinities.

In order to establish that this transition was extremely challenging with the use of spherical cation representation, we compared our work with identical simulations using that model. We found that the relative binding affinities obtained from those simulations were approximately 10 kcal/mol away from the experimental values. Besides, we did not find the transition from Ca^{2+} to Mg^{2+} as being a reversible reaction. This cannot be true physically, because the transition is an equilibrium phenomenon. The application of the multisite cation model, on the other hand, showed advantage at both these levels; the relative binding energy matched precisely with the experimental value and the process was seen to be reversible.

Thus, we found that for all systems: water and ion–binding proteins, the multisite model could be successfully implemented in the simulations and produced accurate results. We showed that with the use of this model, the ion mediated mechanism could be accurately demonstrated both qualitatively and quantitatively in molecular simulations. The model is convenient to use, computationally inexpensive and does not require the use of any sophisticated potential or external tethers between the coordinating atoms and the ion.

Chapter 6

Discussion and Conclusions

This thesis is divided into two main projects: Structure-function relationship in BK_{Ca} channels and the development of the Multisite cation model. Both the projects were successfully completed and are summarized below.

6.1 Structure-functional relationship of BK_{Ca} channels

The aim of this project was to understand the effect of the D369G mutation on the structure and function of the BK_{Ca} channel protein. It was known that a single residue mutation, D369G, in the channel protein significantly alters the function of the protein. We perform computational studies on the channel protein and find that the mutation is able to enhance the Ca^{2+} sensitivity of the channel by reducing the dynamics of the channel protein. The work was divided into two parts where the first part focussed on a small segment of the channel protein, called AC region and the second part was focussed on the entire subunit of the channel, comprising of domains RCK1 and RCK2. The results from the computational simulations were compared with the electrophysiological findings from the Cui lab to find correspondence with the mutagenesis experiments. The results obtained from each part are summarized below:

6.1.1 Simulations on the AC region

To investigate the effect of the D369G mutation, we concentrated on an important segment of the channel protein, the AC region, which was known to play an important role in the calcium related function of the channel. When we started this work, the crystal structure of the protein was not available so we developed a homology model of the protein and performed molecular dynamics simulations (80 ns) on the wildtype and mutant proteins to compare the structure and dynamics of the two proteins. In addition to the single trajectory simulations of the proteins, we performed multiple trajectory runs (0.5 μs) , each starting with different velocities, to sample a larger conformational space and thereby provide statistical significance the observed properties. The results obtained from the simulations were as follows:

- ◇ RMSF analysis showed that the D369G mutation, uniquely experienced a loss in its backbone dynamics, not observed in either the wildtype or the control proteins.
- ◇ The loss in dynamics was localized in helices αA and αB which were distantly located from the mutation site.
- ◇ Analysis of the complex motion of the proteins by PCA showed that the most significant motion of the wildtype protein had a large magnitude indicating that it was very dynamic.
- ◇ PCA on the mutant protein showed that the most significant motion of this protein was quiet less than that seen in the wildtype protein.
- ◇ The results also showed that in the wildtype protein, the motion of the DRDD loop is correlated with that of the αA and αB helices which is lost after the mutation is made.
- ◇ Comparison with the mutagenesis experiments from the Cui lab showed that the helices αA and αB, which experienced loss in flexibility upon D369G mutation, were also important for the calcium activation of the channel.

Thus, we concluded that the effect of the D369G mutation on the AC region was to reduce its backbone flexibility. Presence of Gly at location 369 in the DRDD

loop, disrupts the correlation between the DRDD loop and αA and αB helices. This renders the two helices rigid. Alanine scanning findings from the Cui lab showed that the residues in motifs αA, αB, DRDD loop and βC participate in the calcium activation mechanism. Comparison of our results with these findings shows that the D369G mutation specifically targets the motifs which contain functionally important residues.

6.1.2 Simulations on the full subunit

The results from the previous section demonstrated the effect of D369G mutation on the AC region. AC region was located in the N–terminus of the RCK1 domain. In this section, we expanded our region of interest and investigated the effect of the mutation on the entire subunit of the cytoplasmic domain of the channel. This region comprised of both the RCK1 and RCK2 domains and the results demonstrated the molecular mechanism of the D369G mutation effect on a larger scale. The starting structure used in this simulation was taken from the released recently crystal structure of the BK_{Ca} channel protein and multiple simulations were run (0.6 μs) on both the wildtype and the mutant protein to characterize the conformational changes. The results from these simulations were:

- ◇ PCA of the wildtype simulations showed that the protein was very dynamic and the conformational states were divided into two major conformations, a closed state (W1) and an open state (W2).
- ◇ The difference between the closed and open state was the relative distance between the DRDD loop and the M513 loop which changed from 5 Å to 10 Å.
- ◇ Hydrogen bond analysis showed that the closed state was stayed in that conformation by virtue of the salt-bridge between Arg368 with Glu535. The loss of interaction between these two residues eventually gives rise to the open state.
- ◇ D369G mutation introduces a new set of interaction for Arg368 and it now interacts with Asp367.

- ◇ Thus, the mutant protein occurs in a conformation which lies between the closed and open states of the wildtype protein. The distance between the DRDD loop and the M513 loop in this conformation stays at 7 Å.
- ◇ PCA shows that the mutation makes the protein less dynamic and renders the protein in fixed conformation.

Thus, we concluded that the D369G mutation indeed reduces the dynamics of the BK_{Ca} channel protein. These changes are caused by the secondary structural changes introduced in the protein backbone due to the presence of Gly instead of an Asp. In fixing the protein to a rigid conformation, it appears that the D369G mutation might be mimicking the effect which a calcium binding would produce. We say this because based on recent studies, it is believed that the region enclosed by the DRDD loop, M513 loop and the βG strand might form a calcium binding site. Our simulations show that this site is very dynamic in the wildtype protein with the movement of the DRDD loop governing the dynamics. Since the DRDD loop contains a calcium coordinating residue, Asp367, the binding of calcium will restrict its motion and will bring the protein to a rigid conformation. The D369G mutation is able to achieve that even without the need of calcium ion. Thus, we think that the mutation might be making the task easier and more efficient.

6.1.3 Discussion

The results from both the simulations show that the overall effect of the D369G mutation is to reduce the dynamics of the protein backbone. We observe this in the AC region simulations and also in the simulations of the entire subunit. The simulations in the second part explain the mechanism on a larger scale and we are able to understand how the mutation might be related to the calcium related function of the channel. Simulations on the AC region also talk about that when we compare our results with the alanine scanning results. It is very encouraging to find that even when we ran the simulations on the AC region, we were able to identify that the dynamical effect of the mutation. In the absence of the entire calcium binding site, it was however, difficult to get a structural explanation of the mechanism. The simulations on the full subunit become advantageous in that respect and provide a

bigger picture. They further confirm that AC region indeed is an important region for the function of the channel because it is this region in the wildtype protein which is very dynamic and is rendered rigid by the mutation.

6.1.4 Future Work

These results though explain the molecular mechanism of the D369G mutation on the BK_{Ca} channels but they also open doors to a lot of questions such as, 'What happens when the calcium actually binds to the D367/E535 binding site?', and 'Does the effect of the mutation remains the same?', and so on. The future work on this project therefore, would be to run simulations with the calcium bound to the binding site and then perform the mutation. This would require development of a structure of the channel with the calcium ion bound to the binding site. Thereafter, it would be interesting to investigate how this effect is eventually transferred to the pore of the channel and causes an upregulation in its activity.

6.2 Multisite Cation Model

The second project of this thesis was aimed at developing a new model of cations which could be efficiently used in the molecular mechanics simulations. Currently, the cations are represented as a sphere in molecular mechanics force fields that is the same as the representation of other atoms. Cations, however, form a coordinate bond with its ligands and this allows the ligands to arrange in a specific geometry around the cation. The geometry is determined based on the charge and size of the cation. There is no unique way of representation of coordinate bonds in molecular mechanics force fields, so they are calculated by the non-bonded potentials. In order to achieve accuracy in the geometry of the coordinating ligands careful parametrization of the cation and the coordinating atoms is required. In this project, we developed a new model of the cations that implicitly contains the geometry of the coordinating atoms. The development of the model involved systematic refinement of the force field parameters such that the experimental energies were reproduced. The model was tested

in molecular dynamics simulations with water and proteins and was seen to accurately reproduce thermodynamic as well as structural properties of the coordinating ligands. The key results from this project can be summarized as:

- ◇ The use of multisite cation model in ion–water molecular simulations showed that the experimentally observed values of the coordination number and the radial distribution function were reproduced.
- ◇ The simulations showed accurate represented the geometry of both the first and second solvation shell of the cation.
- ◇ Simulations of the cations with the ion binding protein demonstrated that the conformation of the binding site was maintained and it matched the experimentally obtained values. None of the previous simulation studies which used the spherical representation of the cation, were able to accurately represent the coordination partners.
- ◇ The use of the multisite model was able to demonstrate, with high precision, the preferential binding of Ca^{2+} ions over Mg^{2+} ions in a calcium binding protein.

Thus, we conclude that the multisite cation model provides accurate representation of ion–protein interactions in molecular mechanics simulations. The use of the model provide particular advantage in mechanisms involving selectivity to a specific ion.

6.2.1 Discussion

Our results showed that the use the multisite cation model is able to account for some of the induced polarization effects as it is able to predict the geometry of the second coordination shell as well. Earlier ion–protein simulations had seen some erroneous ion coordinations, attributed to the neglect of these terms. Our simulations showed the use of multisite cation model, got rid of the earlier artifacts and yielded accurate coordination.

Further, the results from the selectivity of Ca^{2+}/Mg^{2+} ions show that the use of multsite cation model is particularly useful for these complex mechanisms which were

a challenge to demonstrate earlier. Ca^{2+} and Mg^{2+} are very similar to each other as they have the same net charge and almost similar radii. This difference is not easily picked up when both the ions are represented as sphere. In our model the coordinating geometry of the cation is implicitly defined. This allows the simulations to differentiate between the two and yield accurate energy values.

6.2.2 Future Work

The model shown here has been parameterized for two cations, calcium and magnesium and in the AMBER force fields. Since the model has shown successful results, it will be worthwhile to extend it to other force fields and other cations. In our work we have shown that the model can be used for identifying the selectivity between calcium and magnesium ions. This property can be used for other ions such as sodium and potassium which will be an interesting area of study for demonstrating the selectivity of an ion channel pore.

References

J Aaqvist. Ion-water interaction potentials derived from free energy perturbation simulations. *The Journal of Physical Chemistry*, 94(21):8021–8024, 1990.

J.P Adelman, K.Z Shen, M.P Kavanaugh, R.A Warren, Y.N Wu, A Lagrutta, C.T Bond, and R.A North. Calcium-activated potassium channels expressed from cloned complementary dnas. *Neuron*, 9(2):209–16, 1992.

P Ahlstroem, O Teleman, J Koerdel, S Forsen, and B Joensson. A molecular dynamics simulation of bovine Calbindin D9k. molecular structure and dynamics. *Biochemistry*, 28(8):3205, 1989.

M Akke, S Forsen, and W.J Chazin. Molecular basis for co-operativity in Ca2+ binding to Calbindin D9k. 1h nuclear magnetic resonance studies of (Cd2+)1-bovine Calbindin D9k. *J Mol Biol*, 220(1):173–89, 1991.

M Akke, S Forsen, and W.J Chazin. Solution structure of (Cd2+)1II-Calbindin D9k reveals details of the stepwise structural changes along the apo,(Ca2+)1II;(Ca2+)2I,II binding pathway. *J Mol Biol*, 252(1):102, 1995.

B Alder and T Wainwright. Studies in molecular dynamics. i. general method. *The Journal of chemical physics*, 31(2):459–466, 1959.

B Alder and T Wainwright. Studies in molecular dynamics. ii. behavior of a small number of elastic spheres. *The Journal of chemical physics*, 33(5):1439–1451, 1960.

R N.N. Alexandrov. Fast protein fold recognition via sequence to structure alignment and contact capacity potentials. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 1996.

D Allouche, J Parello, and Y.H Sanejouand. Ca2+/Mg2+ exchange in parvalbumin and other EF-hand proteins. a theoretical study. *J Mol Biol*, 285(2):857–73, 1999.

S Altschul, W Gish, W Miller, E Myers, and D Lipman. Basic local alignment search tool. *J Mol Biol*, 215(3):403–10, 1990.

S Altschul, T Madden, A Schaffer, J Zhang, Z Zhang, W Miller, and D Lipman. Gapped blast and psi-blast: a new generation of protein database search programs. *Nucleic Acids Res*, 25(17):3389–402, 1997.

K.M Anderson, P.M Odell, P.W Wilson, and W.B Kannel. Cardiovascular disease risk profiles. *Am Heart J*, 121(1 Pt 2):293–8, 1991.

M Andersson, A Malmendal, S Linse, I Ivarsson, S Forsen, and L.A Svensson. Structural basis for the negative allostery between ca(2+)- and mg(2+)-binding in the intracellular ca(2+)-receptor Calbindin D9k. *Protein Sci*, 6(6):1139–47, 1997.

T Andrea, W Swope, and H Andersen. The role of long ranged forces in determining the structure and properties of liquid water. *The Journal of chemical physics*, 79 (9):4576, 1983.

J Aqvist and A Warshel. Free energy relationships in metalloenzyme-catalyzed reactions. calculations of the effects of metal ion substitutions in staphylococcal nuclease. *J Am Chem Soc*, 112(8):2860, 1990.

J Aqvist and A Warshel. Computer simulation of the initial proton transfer step in human carbonic anhydrase i. *J Mol Biol*, 224(1):7, 1992.

D Asthagiri, P.D Dixit, S Merchant, M.E Paulaitis, L.R Pratt, S.B Rempe, and S Varma. Ion selectivity from local configurations of ligands in solutions and ion channels. *Chemical Physics Letters*, 485(1-3):1–7, Jan 2010.

N.S Atkinson, G.A Robertson, and B Ganetzky. A component of calcium-activated potassium channels encoded by the drosophila slo locus. *Science*, 253(5019):551–5, 1991.

T Attwood, P Bradley, D Flower, A Gaulton, N Maudling, A Mitchell, G Moulton, A Nordle, K Paine, P Ta ylor, A Uddin, and C Zygouri. Prints and its automatic supplement, preprints. *Nucleic Acids Res*, 31(1):400–2, 2003.

Y. S. Badyal, A. C. Barnes, G. J. Cuello, and J. M. Simonson. Understanding the effects of concentration on the solvation structure of Ca2+ in aqueous solution. ii: insights into longer range order from neutron diffraction isotope substitution. *The Journal of Physical Chemistry A*, 108(52):11819–11827, 2004.

A Bairoch, B Boeckmann, S Ferro, and E Gasteiger. Swiss-prot: juggling between evolution and stability. *Brief Bioinform*, 5(1):39–55, 2004.

D Baker and A Sali. Protein structure prediction and structural genomics. *Science*, 294(5540):93–6, 2001.

N Baker, D Sept, S Joseph, M Holst, and J McCammon. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A*, 98(18):10037–41, 2001a.

N.A Baker, D Sept, S Joseph, M.J Holst, and J.A McCammon. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A*, 98(18):10037–41, 2001b.

L Bao, A.M Rapin, E.C Holmstrand, and D.H Cox. Elimination of the BK(ca) channel's high-affinity ca(2+) sensitivity. *J Gen Physiol*, 120(2):173–89, 2002.

L Bao, C Kaldany, E.C Holmstrand, and D.H Cox. Mapping the BKca channel's "Ca2+ bowl": side-chains essential for Ca2+ sensing. *J Gen Physiol*, 123(5):475–89, 2004.

P Bates, L Kelley, R MacCallum, and M Sternberg. Enhancement of protein modeling by human intervention in applying the automatic programs 3d-jigsaw and 3d-pssm. *Proteins*, Suppl 5:39–46, 2001.

D Benson, I Karsch-Mizrachi, D Lipman, J Ostell, and E Sayers. Genbank. *Nucleic Acids Res*, 37(Database issue):D26–31, 2009.

H Berendsen, J Postma, W van Gunsteren, A DiNola, and J Haak. Molecular dynamics with coupling to an external bath. *The Journal of chemical physics*, 81(8): 3684, 1984.

H Berendsen, J Grigera, and T Straatsma. The missing term in effective pair potentials. *The Journal of Physical Chemistry*, 91:6269–6271, 1987. The missing term in effective pair potentials.

H.J.C. Berendsen, J.P.M. Postma, W.F.van Gunsteren, and Jan Hermans. Interaction models for water in relation to protein hydration. *Intermolecular forces*, pages 331–342, 1981.

H Berman, J Westbrook, Z Feng, G Gilliland, T Bhat, H Weissig, I Shindyalov, and P Bourne. The protein data bank. *Nucleic Acids Res*, 28(1):235–42, 2000.

A Bonvin. Flexible protein-protein docking. *Curr Opin Struct Biol*, 16(2):194–200, 2006.

M Bradley, P Chivers, and N Baker. Molecular dynamics simulation of the escherichia coli NikR protein: Equilibrium conformational fluctuations reveal interdomain allosteric communication pathways. *J Mol Biol*, 378(5):1155, 2008.

R Brenner, G.J Perez, A.D Bonev, D.M Eckman, J.C Kosek, S.W Wiler, A.J Patterson, M.T Nelson, and R.W Aldrich. Vasoregulation by the beta1 subunit of the calcium-activated potassium channel. *Nature*, 407(6806):870–6, 2000.

R Brenner, Q Chen, A Vilaythong, G Toney, J Noebels, and R Aldrich. Bk channel [beta]4 subunit reduces dentate gyrus excitability and protects against temporal lobe seizures. *Nat Neurosci*, 8(12):1752–1759, 2005. 10.1038/nn1573.

B Brooks, C Brooks, A Mackerell, L Nilsson, R Petrella, B Roux, Y Won, G Archontis, C Bartels, S Boresch, A Caflisch, L Caves, Q Cui, A Dinner, M Feig, S Fischer, J Gao, M Hodoscek, W Im, K Kuczera, T Lazaridis, J Ma, V Ovchinnikov, E Paci, R Pastor, C Post, J Pu, M Schaefer, B Tidor, R Venable, H Woodcock, X Wu, W Yang, D York, and M Karplus. CHARMM: the biomolecular simulation program. *J Comput Chem*, 30(10):1545–614, 2009.

K Bryson, L McGuffin, R Marsden, J Ward, J Sodhi, and D Jones. Protein structure prediction servers at university college london. *Nucleic Acids Res*, 33:36–8, 2005.

A Butler, S Tsunoda, D.P McCobb, A Wei, and L Salkoff. mslo, a complex mouse gene encoding "maxi" calcium-activated potassium channels. *Science*, 261(5118): 221–4, 1993.

V Calderone. Large-conductance, ca(2+)-activated k(+) channels: function, pharmacology and drugs. *Curr Med Chem*, 9(14):1385–95, 2002.

J Caldwell and P Kollman. Structure and properties of neat liquids using nonadditive molecular dynamics: Water, methanol, and n-methylacetamide. *The Journal of Physical Chemistry*, 99:6208–6219, 1995.

M.S Cates, M.L Teodoro, and G.N Phillips. Molecular mechanisms of calcium and magnesium binding to parvalbumin. *Biophys J*, 82(3):1133–46, 2002.

J Cavanagh and M Akke. May the driving force be with you–whatever it is. *Nat Struct Biol*, 7(1):11–3, 2000.

L.S Caves, J.D Evanseck, and M Karplus. Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. *Protein Sci*, 7(3): 649–66, 1998.

T Cheatham and M Young. Molecular dynamics simulation of nucleic acids: successes, limitations, and promise. *Biopolymers*, 56(4):232–56, 2000.

J Cheng, A Randall, M Sweredoski, and P Baldi. Scratch: a protein structure and structural feature prediction server. *Nucleic Acids Res*, 33:72–6, 2005.

K.C Chou and L Carlacci. Simulated annealing approach to the study of protein structures. *Protein Eng*, 4(6):661–7, 1991.

C Cole, J Barber, and G Barton. The jpred 3 secondary structure prediction server. *Nucleic Acids Res*, 36:197–201, 2008.

A Cooper and D.T Dryden. Allostery without conformational change. a plausible model. *Eur Biophys J*, 11(2):103–9, 1984.

D.H Cox, J Cui, and R.W Aldrich. Allosteric gating of a large conductance ca-activated K+ channel. *J Gen Physiol*, 110(3):257–81, 1997.

J Cui and R.W Aldrich. Allosteric linkage between voltage and ca(2+)-dependent activation of BK-type mslo1 k(+) channels. *Biochemistry*, 39(50):15612–9, 2000.

J Cui, D.H Cox, and R.W Aldrich. Intrinsic voltage dependence and Ca2+ regulation of mslo large conductance ca-activated K+ channels. *J Gen Physiol*, 109(5):647–73, 1997.

A.C.R da Silva and F.C Reinach. Calcium binding induces conformational changes in muscle regulatory proteins. *Trends Biochem Sci*, 16:53, 1991.

T Darden, D Pearlman, and L.G Pedersen. Ionic charging free energies: Spherical versus periodic boundary conditions. *The Journal of chemical physics*, 109(24): 10921, 1998.

A Davidson. A folding space odyssey. *Proc Natl Acad Sci U S A*, 105(8):2759–60, 2008.

I Davis and D Baker. Rosetta ligand docking with full ligand and receptor flexibility. *J Mol Biol*, 385(2):381–92, 2009.

A de Vries, I Chandrasekhar, W van Gunsteren, and P Hunenberger. Molecular dynamics simulations of phospholipid bilayers: influence of artificial periodicity, system size, and simulation time. *The Journal of Physical Chemistry B*, 109: 11643–11652, 2005.

K Diraviyam, R Stahelin, W Cho, and D Murray. Computer modeling of the membrane interaction of fyve domains. *J Mol Biol*, 328(3):721–36, 2003.

C Dominguez, R Boelens, and A Bonvin. Haddock: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc*, 125(7): 1731–7, 2003.

O.A Donini and P.A Kollman. Calculation and prediction of binding free energies for the matrix metalloproteinases. *J Med Chem*, 43(22):4180–8, 2000.

W Du, J.F Bautista, H Yang, A Diez-Sampedro, S.A You, L Wang, P Kotagal, H.O Luders, J Shi, J Cui, G.B Richerson, and Q.K Wang. Calcium-sensitive potassium channelopathy in human epilepsy and paroxysmal movement disorder. *Nat Genet*, 37(7):733–8, 2005.

Y Duan, C Wu, S Chowdhury, M Lee, G Xiong, W Zhang, R Yang, P Cieplak, R Luo, T Lee, J Caldwell, J Wang, and P Kollman. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of Computational Chemistry*, 24(16):1999–2012, 2003.

H Ebel and T Gunther. Magnesium metabolism: a review. *J Clin Chem Clin Biochem*, 18(5):257–70, 1980.

S Eddy. Profile hidden markov models. *Bioinformatics*, 14(9):755–63, 1998.

T Elkins, B Ganetzky, and C.F Wu. A drosophila mutation that eliminates a calcium-dependent potassium current. *Proc Natl Acad Sci U S A*, 83(21):8415–9, 1986.

E S Louise Faber and Pankaj Sah. Ca2+-activated K+ (bk) channel inactivation contributes to spike broadening during repetitive firing in the rat lateral amygdala. *The Journal of Physiology*, 552(2):483–497, 2003.

J.J Falke, S.K Drake, A.L Hazard, and O.B Peersen. Molecular tuning of ion binding to calcium signaling proteins. *Q Rev Biophys*, 27(3):219–90, 1994.

R Finn, J Tate, J Mistry, P Coggill, S Sammut, H Hotz, G Ceric, K Forslund, S Eddy, E Sonnhammer, and A Bateman. The pfam protein families database. *Nucleic Acids Res*, 36(Database issue):D281–8, 2008.

A.A Fodor and R.W Aldrich. Statistical limits to the identification of ion channel domains by sequence similarity. *J Gen Physiol*, 127(6):755–66, 2006.

J.D Forman-Kay. The 'dynamics' in the thermodynamics of binding. *Nat Struct Biol*, 6(12):1086–7, 1999.

M Formaneck and Q Cui. The use of a generalized born model for the analysis of protein conformational transitions: A comparative study with explicit solvent simulations for chemotaxis y protein (chey). *J. Comput. Chem.*, 27(16):1923–1943, 2006.

L.R Forrest, A Kukol, I.T Arkin, D.P Tieleman, and M.S Sansom. Exploring models of the influenza a m2 channel: Md simulations in a phospholipid bilayer. *Biophys J*, 78(1):55–69, 2000.

K.K Frederick, M.S Marlow, K.G Valentine, and A.J Wand. Conformational entropy in molecular recognition by proteins. *Nature*, 448(7151):325–9, 2007.

R Friesner, J Banks, R Murphy, T Halgren, J Klicic, D Mainz, M Repasky, E Knoll, M Shelley, J Perry, D Shaw, P Francis, and P Shenkin. Glide: a new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. *J Med Chem*, 47(7):1739–49, 2004.

J.L Fulton, S.M Heald, Y.S Badyal, and J.M Simonson. Understanding the effects of concentration on the solvation structure of Ca2+ in aqueous solution. i:â the perspective on local structure from exafs and xanes. *The Journal of Physical Chemistry A*, 107(23):4688, 2003.

A.M Gaspar, M Alves Marques, M.I Cabaço, M.I de Barros Marques, T Buslaps, and V Honkimaki. X-ray diffraction investigations of concentrated aqueous solutions of calcium halides. *Journal of Molecular Liquids*, 110(1-3):15, 2004.

E Gasteiger, A Gattiker, C Hoogland, I Ivanyi, R Appel, and A Bairoch. Expasy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*, 31(13):3784–8, 2003.

D Goodsell and A Olson. Automated docking of substrates to proteins by simulated annealing. *Proteins*, 8(3):195–202, 1990.

J Gray, S Moughon, C Wang, O Schueler-Furman, B Kuhlman, C Rohl, and D Baker. Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J Mol Biol*, 331(1):281–99, 2003.

S.M Green and D Shortle. Patterns of nonadditivity between pairs of stability mutations in staphylococcal nuclease. *Biochemistry*, 32(38):10131–9, 1993.

N Gresh, G.A Cisneros, T.A Darden, and J.P Piquemal. Anisotropic, polarizable molecular mechanics studies of inter- and intramolecular interactions and ligand-macromolecule complexes. a bottom-up strategy. *J Chem Theory Comput*, 3(6): 1960–1986, 2007.

A Grossfield, S.E Feller, and M.C Pitman. Contribution of omega-3 fatty acids to the thermodynamics of membrane protein solvation. *J Phys Chem B Condens Matter Mater Surf Interfaces Biophys*, 110(18):8907–9, 2006.

A Grottesi, C Domene, B Hall, and M Sansom. Conformational dynamics of m2 helices in kirbac channels: helix flexibility in relation to gating via molecular dynamics simulations. *Biochemistry*, 44(44):14586–94, 2005.

K Gunasekaran, B Ma, and R Nussinov. Is allostery an intrinsic property of all dynamic proteins? *Proteins*, 57(3):433–43, 2004.

T Halgren, R Murphy, R Friesner, H Beard, L Frye, W Pollard, and J Banks. Glide: a new approach for rapid, accurate docking and scoring. 2. enrichment factors in database screening. *J Med Chem*, 47(7):1750–9, 2004.

C Hardin, T Pogorelov, and Z Luthey-Schulten. Ab initio protein structure prediction. *Curr Opin Struct Biol*, 12(2):176–81, 2002.

H Heller, M Schaefer, and K Schulten. Molecular dynamics simulation of a bilayer of 200 lipids in the gel and in the liquid crystal phase. *The Journal of Physical Chemistry*, 97:8343–8360, 1993.

C Hemmingsen, M Staun, and K Olgaard. Effects of magnesium on renal and intestinal Calbindin-d. *Miner Electrolyte Metab*, 20(5):265–73, 1994.

B Hess, C Kutzner, D van der Spoel, and E Lindahl. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447, Mar 2008.

N.A Hewish, G.W Neilson, and J.E Enderby. Environment of Ca2+ ions in aqueous solvent. *Nature*, 297(5862):138, 1982.

A Horovitz and A.R Fersht. Strategy for analysing the co-operativity of intramolecular interactions in peptides and proteins. *J Mol Biol*, 214(3):613–7, 1990.

F.T Horrigan and R.W Aldrich. Coupling between voltage sensor activation, Ca2+ binding and channel opening in large conductance (bk) potassium channels. *J Gen Physiol*, 120(3):267–305, 2002.

L Hu, J Shi, Z Ma, G Krishnamoorthy, F Sieling, G Zhang, F.T Horrigan, and J Cui. Participation of the s4 voltage sensor in the Mg2+-dependent activation of large conductance (bk) K+ channels. *Proc Natl Acad Sci U S A*, 100(18):10488–93, 2003.

Y Hui, J Shirin, M Shanthini, and C Peter. Activation of large-conductance Ca2+-activated K+ channels depresses basal synaptic transmission in the hippocampal ca1 area in app (swe/ind) tgcrnd8 mice. *Neurobiology of aging*, 31(4):591, 2010.

N Hulo, A Bairoch, V Bulliard, L Cerutti, B Cuche, E de Castro, C Lachaize, P Langendijk-Genevaux, and C Sigrist. The 20 years of prosite. *Nucleic Acids Res*, 36(Database issue):D245–9, 2008.

G Hummer, L.R Pratt, and A.E Garcia. Free energy of ionic hydration. *The Journal of Physical Chemistry*, 100(4):1206, 1996.

W Humphrey, A Dalke, and K Schulten. Vmd: visual molecular dynamics. *J Mol Graph*, 14(1):33–8, 27–8, 1996.

S Hunter, R Apweiler, T Attwood, A Bairoch, A Bateman, D Binns, P Bork, U Das, L Daugherty, L Duquenne, R Finn, J Gough, D Haft, N Hulo, D Kahn, E Kelly, A Laugraud, I Letunic, D Lonsdale, R Lopez, M Madera, J Maslen, C McAnulla, J McDowall, J Mistry, A Mitchell, N Mulder, D Natale, C Orengo, A Quinn, J Selengut, C Sigrist, M Thimma, P Thomas, F Valentin, D Wilson, C Wu, and C Yeats. Interpro: the integrative protein signature database. *Nucleic Acids Res*, 37(Database issue):D211–5, 2009.

T Ichiye and M Karplus. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins*, 11(3):205–17, 1991.

C Illingworth, G Morris, K Parkes, C Snell, and C Reynolds. Assessing the role of polarization in docking. *J Phys Chem A*, 112(47):12157–63, 2008a.

C Illingworth, K Parkes, C Snell, G Ferenczy, and C Reynolds. Toward a consistent treatment of polarization in model qm/mm calculations. *J Phys Chem A*, 112(47): 12151–6, 2008b.

C Illingworth, K Parkes, C Snell, P Mullineaux, and C Reynolds. Criteria for confirming sequence periodicity identified by fourier transform analysis: application to gcr2, a candidate plant gpcr? *Biophys Chem*, 133(1-3):28–35, 2008c.

C Illingworth, K Parkes, C Snell, and C Reynolds. Quantitative measurement of protease ligand conformation. *J Comput Aided Mol Des*, 22(2):105–9, 2008d.

J Irwin. Community benchmarks for virtual screening. *J Comput Aided Mol Des*, 22 (3-4):193–9, 2008.

I Ivanov, B.R Chapados, J.A McCammon, and J.A Tainer. Proliferating cell nuclear antigen loaded onto double-stranded dna: dynamics, minor groove interactions and functional implications. *Nucleic Acids Res*, 34(20):6023–33, 2006.

M Jacobson, R Friesner, Z Xiang, and B Honig. On the role of the crystal environment in determining protein side-chain conformations. *J Mol Biol*, 320(3):597–608, 2002a.

M.P Jacobson, R.A Friesner, Z Xiang, and B Honig. On the role of the crystal environment in determining protein side-chain conformations. *J Mol Biol*, 320(3): 597–608, 2002b.

M.P Jacobson, D.L Pincus, C.S Rapp, T.J Day, B Honig, D.E Shaw, and R.A Friesner. A hierarchical approach to all-atom protein loop prediction. *Proteins*, 55(2):351–67, 2004.

A Jain. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J Med Chem*, 46(4):499–511, 2003.

A Jain. Scoring functions for protein-ligand docking. *Curr Protein Pept Sci*, 7(5): 407–20, 2006.

A Jain. Surflex-dock 2.1: robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. *J Comput Aided Mol Des*, 21(5):281–306, 2007.

A Jain. Effects of protein conformation in docking: improved pose prediction through protein pocket adaptation. *J Comput Aided Mol Des*, 23(6):355–74, 2009.

F Jalilehvand, D Spngberg, P Lindqvist-Reis, K Hermansson, I Persson, and M Sandstrm. Hydration of the calcium ion. an exafs, large-angle x-ray scattering, and molecular dynamics simulation study. *Journal of the American Chemical Society*, 123(3):431–441, 2001.

J Jiang, K Prasad, E.M Lafer, and R Sousa. Structural basis of interdomain communication in the hsc70 chaperone. *Mol Cell*, 20(4):513–24, 2005.

Y Jiang, A Pico, M Cadene, B.T Chait, and R Mackinnon. Structure of the rck domain from the e. coli K+ channel and demonstration of its presence in the human BK channel. *Neuron*, 29(3):593–601, 2001.

Y Jiang, A Lee, J Chen, M Cadene, B.T Chait, and R Mackinnon. Crystal structure and mechanism of a calcium-gated potassium channel. *Nature*, 417(6888):515–22, 2002a.

Y Jiang, A Lee, J Chen, M Cadene, B.T Chait, and R Mackinnon. The open pore conformation of potassium channels. *Nature*, 417(6888):523–6, 2002b.

Y Jiang, A Lee, J Chen, V Ruta, M Cadene, B.T Chait, and R Mackinnon. X-ray structure of a voltage-dependent K+ channel. *Nature*, 423(6935):33–41, 2003.

W.J Joiner, M.D Tang, L.Y Wang, S.I Dworetzky, C.G Boissard, L Gan, V.K Gribkoff, and L.K Kaczmarek. Formation of intermediate-conductance calcium-activated potassium channels by interaction of slack and slo subunits. *Nat Neurosci*, 1(6):462–9, 1998.

G Jones, P Willett, and R Glen. A genetic algorithm for flexible molecular overlay and pharmacophore elucidation. *J Comput Aided Mol Des*, 9(6):532–49, 1995.

G Jones, P Willett, R Glen, A Leach, and R Taylor. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol*, 267(3):727–48, 1997.

W Jorgensen and J Madura. Temperature and size dependence for monte carlo simulations of tip4p water. *Molecular Physics: An International Journal at the Interface Between Chemistry and Physics*, 56(6):1381 – 1392, 1985.

W Jorgensen, Jayaraman Chandrasekhar, J Madura, R Impey, and M Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.

W.L Jorgensen, D.S Maxwell, and J Tirado-Rives. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc*, 118(45):11225, 1996.

I Joung and T Cheatham. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *The Journal of Physical Chemistry B*, 112(30):9020–9041, Jul 2008.

W Kabsch. A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Cryst.*, A(34):827–828, 1978.

G Kaczorowski, H Knaus, and R Leonard. High-conductance calcium-activated potassium channels; structure, pharmacology, and function. *Journal of bioenergetics and biomembranes*, Jan 1996.

A Katz, J Glusker, S Beebe, and C Bock. Calcium ion coordination: a comparison with that of beryllium, magnesium, and zinc. *J Am Chem Soc*, 118(24):5752–5763, Jan 1996.

L Kelley, R MacCallum, and M Sternberg. Enhanced genome annotation using structural profiles in the program 3d-pssm. *J Mol Biol*, 299(2):499–520, 2000.

D Kern and E.R Zuiderweg. The role of dynamics in allosteric regulation. *Curr Opin Struct Biol*, 13(6):748–57, 2003.

H.J Kim, H.H Lim, S.H Rho, L Bao, J.H Lee, D.H Cox, H Kim do, and C.S Park. Modulation of the conductance-voltage relationship of the BK ca channel by mutations at the putative flexible interface between two rck domains. *Biophys J*, 94 (2):446–56, 2008.

J Koerdel and O Teleman. Backbone dynamics of Calbindin D9k: comparison of molecular dynamics simulations and nitrogen-15 nmr relaxation measurements. *J Am Chem Soc*, 114(12):4934, 1992.

M Kontoyianni, P Madhav, E Suchanek, and W Seibel. Theoretical and practical considerations in virtual screening: a beaten field? *Curr Med Chem*, 15(2):107–16, 2008.

R Kretsinger, R Ison, and S Hovmoller. Prediction of protein structure. *Methods Enzymol*, 383:1–27, 2004.

G Krishnamoorthy, J Shi, D Sept, and J Cui. The nh2 terminus of rck1 domain regulates Ca2+-dependent BK(ca) channel gating. *J Gen Physiol*, 126(3):227–41, 2005.

A Krogh, B Larsson, G von Heijne, and E Sonnhammer. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J Mol Biol*, 305(3):567–80, 2001.

M Larkin, G Blackshields, N Brown, R Chenna, P McGettigan, H McWilliam, F Valentin, I Wallace, A Wilm, R Lopez, J Thompson, T Gibson, and D Higgins. Clustal w and clustal x version 2.0. *Bioinformatics*, 23(21):2947–8, 2007.

R. A. Laskowski, M. W. MacArthur, D. S. Moss, and J. M. Thornton. *PROCHECK*: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26(2):283–291, Apr 1993.

F. Laumonnier, S. Roger, P. Gurin, F. Molinari, R. M'Rad, D. Cahard, A. Belhadj, M. Halayem, A. M. Persico, M. Elia, V. Romano, S. Holbert, C. Andres, H. Chaabouni, L. Colleaux, J. Constant, J. . Le Guennec, and S. Briault. Association of a functional deficit of the BKca channel, a synaptic regulator of neuronal excitability, with autism and mental retardation. *American Journal of Psychiatry*, 163(9):1622–1629, 2006.

M Lawrence and P Davis. Clix: a search algorithm for finding novel ligands capable of binding proteins of known three-dimensional structure. *Proteins*, 12(1):31–41, 1992.

A Leach, B Shoichet, and C Peishoff. Prediction of protein-ligand interactions. docking and scoring: successes and gaps. *J Med Chem*, 49(20):5851–5, 2006.

A. R. Leach. *Molecular Modelling: principles and applications*. Prentice Hall, New York, second edition, 2001.

J Ledoux, M.E Werner, J.E Brayden, and M.T Nelson. Calcium-activated potassium channels and the regulation of vascular tone. *Physiology (Bethesda)*, 21:69–78, 2006.

C Lenfant, A.V Chobanian, D.W Jones, and E.J Roccella. Seventh report of the joint national committee on the prevention, detection, evaluation, and treatment of high blood pressure (JNC 7): resetting the hypertension sails. *Hypertension*, 41 (6):1178–9, 2003.

M Lensink, R Mendez, and S Wodak. Docking and scoring protein complexes: Capri 3rd edition. *Proteins*, 69(4):704–18, 2007.

I Letunic, T Doerks, and P Bork. Smart 6: recent updates and new developments. *Nucleic Acids Res*, 37(Database issue):D229–32, 2009.

V.J LiCata and G.K Ackers. Long-range, small magnitude nonadditivity of mutational effects in proteins. *Biochemistry*, 34(10):3133–9, 1995.

G Licheri, G Piccaluga, and G Pinna. X-ray diffraction study of cabr[sub 2] aqueous solutions. *The Journal of chemical physics*, 63(10):4412, 1975.

G Licheri, G Piccaluga, and G Pinna. X-ray diffraction study of the average solute species in CaCl2 aqueous solutions. *J. Chem. Phys.*, 64, 1976.

F. C. Lightstone, E Schwegler, R. Q. Hood, F Gygi, and G Galli. A first principles molecular dynamics simulation of the hydrated magnesium ion. *Chemical Physics Letters*, 343(5-6):549 – 555, 2001. ISSN 0009-2614.

A Lobley, M Sadowski, and D Jones. pgenthreader and pdomthreader: new methods for improved protein fold recognition and superfamily discrimination. *Bioinformatics*, 25(14):1761–7, 2009.

M Lohn, W Jessner, M Furstenau, M Wellner, V Sorrentino, H Haller, F.C Luft, and M Gollasch. Regulation of calcium sparks and spontaneous transient outward currents by ryr3 in arterial vascular smooth muscle cells. *Circ Res*, 89(11):1051–7, 2001.

S.B Long, E.B Campbell, and R Mackinnon. Crystal structure of a mammalian voltage-dependent shaker family K+ channel. *Science*, 309(5736):897–903, 2005.

R Luthy, J Bowie, and D Eisenberg. Assessment of protein models with three-dimensional profiles. *Nature*, 356(6364):83–5, 1992.

T.P Lybrand, J.A McCammon, and G Wipff. Theoretical calculation of relative binding affinity in host-guest systems. *Proc Natl Acad Sci U S A*, 83(4):833–5, 1986.

S Lyskov and J Gray. The Rosettadock server for local protein-protein docking. *Nucleic Acids Res*, 36:233–8, 2008.

Z Ma, X.J Lou, and F.T Horrigan. Role of charged residues in the s1-s4 voltage sensor of BK channels. *J Gen Physiol*, 127(3):309–28, 2006.

M. S. Madhusudhan, Marc A. Marti-Renom, Narayanan Eswar, Bino John, Ursula Pieper, Rachel Karchin, Min-Yi Shen, and Andrej Sali. Comparative protein structure modeling. In John M. Walker, editor, *The Proteomics Protocols Handbook*, pages 831–860. Humana Press, 2005.

M Mahoney and W Jorgensen. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *The Journal of chemical physics*, 112(20):8910, 2000.

S Marchand and B Roux. Molecular dynamics study of Calbindin D9k in the apo and singly and doubly calcium-loaded states. *Proteins: Structure, Function, and Bioinformatics*, 33(2):265, 1998.

G Markham, J Glusker, and C Bock. The arrangement of first- and second-sphere water molecules in divalent magnesium complexes: results from molecular orbital and density functional theory and from structural crystallography. *The Journal of Physical Chemistry B*, 106:5118–5134, May 2002.

M Marti-Renom, A Stuart, A Fiser, R Sanchez, F Melo, and A Sali. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct*, 29:291–325, 2000.

J McCammon, B Gelin, M Karplus, and P Wolynes. The hinge-bending mode in lysozyme. *Nature*, 262(5566):325–6, 1976.

J McCammon, B Gelin, and M Karplus. Dynamics of folded proteins. *Nature*, 267 (5612):585–90, 1977.

S McGinnis and T Madden. Blast: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res*, 32:20–5, 2004.

A.D McLachlan. Rapid comparison of protein structres. *Acta Cryst*, A(38):871–873, 1982.

O.B McManus and K.L Magleby. Accounting for the ca(2+)-dependent kinetics of single large-conductance ca(2+)-activated K+ channels in rat skeletal muscle. *J Physiol*, 443:739–77, 1991.

P Meera, M Wallner, Z Jiang, and L Toro. A calcium switch for the functional coupling between alpha (hslo) and beta subunits (kv,ca beta) of maxi k channels. *FEBS Lett*, 382(1-2):84–8, Mar 1996.

T Megyes, T. Grosz, T. Radnai, I Bako, and G. Palinkas. Solvation of calcium ion in polar solvents: An x-ray diffraction and ab initio study. *The Journal of Physical Chemistry A*, 108(35):7261, 2004.

E.L Mehler, J.N Kushick, and H Weinstein. Consequences of sequential ca occupancy for the structure and dynamics of calbindin(d9k): Computational simulations and comparison to experimental determinations in solution. *Molecular Simulation*, 10 (2):309 – 334, 1993.

R Mendez, R Leplae, M Lensink, and S Wodak. Assessment of capri predictions in rounds 3-5 shows progress in docking procedures. *Proteins*, 60(2):150–69, 2005.

E Moczydlowski and R Latorre. Gating kinetics of Ca2+-activated K+ channels from rat muscle incorporated into planar lipid bilayers. evidence for two voltage-dependent Ca2+ binding reactions. *J Gen Physiol*, 82(4):511–42, 1983.

K Szebenyi Dm Fau Moffat and K Moffat. The refined structure of vitamin d-dependent calcium-binding protein from bovine intestine. molecular details, ion binding, and implications for the structure of other calcium-binding proteins. *J Biol Chem*, 261(19):8761–77, 1986.

M.M Naor, K.V Nostrand, and C Dellago. Carparrinello molecular dynamics simulation of the calcium ion in liquid water. *Chem Phys Lett*, 369(1-2):159–164, 2003.

J Neely and R Connick. Rate of water exchange from hydrated magnesium ion. *J Am Chem Soc*, 92(11):3476, 1970.

M.T Nelson, H Cheng, M Rubart, L.F Santana, A.D Bonev, H.J Knot, and W.J Lederer. Relaxation of arterial smooth muscle by calcium sparks. *Science*, 270 (5236):633–7, 1995.

A Nicholls, K Sharp, and B Honig. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, 11(4):281–96, 1991.

C Notredame, D Higgins, and J Heringa. T-coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol*, 302(1):205–17, 2000.

R.M Noyes. Thermodynamics of ion hydration as a measure of effective dielectric properties of water. *J Am Chem Soc*, 84(4):513, 1962.

P Oelschlaeger, M Klahn, W.A Beard, S.H Wilson, and A Warshel. Magnesium-cationic dummy atom molecules enhance representation of dna polymerase [beta] in molecular dynamics simulations: Improved accuracy in studies of structural features and mutational effects. *J Mol Biol*, 366(2):687, 2007.

E Owczarek, M Rybicki, and E Hawlicka. Solvation of calcium ions in methanol-water mixtures: Molecular dynamics simulation. *The journal of physical chemistry B*, 111 (51):14271, 2007.

H Pan, J.C Lee, and V.J Hilser. Binding sites in escherichia coli dihydrofolate reductase communicate by modulating the conformational ensemble. *Proc Natl Acad Sci U S A*, 97(22):12020–5, 2000.

Y.P Pang. Successful molecular dynamics simulation of two zinc complexes bridged by a hydroxide in phosphotriesterase using the cationic dummy atom method. *Proteins: Structure, Function and Genetics*, 45(3):183, 2001.

Y.P Pang, K Xu, J.E Yazal, and F.G Prendergast. Successful molecular dynamics simulation of the zinc-bound farnesyltransferase using the cationic dummy atom approach. *Protein Science*, 9(10):1857, 2000.

J.G Park, P.C Sill, E.F Makiyi, A.T Garcia-Sosa, C.B Millard, J.J Schmidt, and Y.P Pang. Serotype-selective, small-molecule inhibitors of the zinc endopeptidase of botulinum neurotoxin serotype a. *Bioorganic and Medicinal Chemistry*, 14(2):395, 2006.

R.W Pastor, B.R Brooks, and A Szabo. An analysis of the accuracy of langevin and molecular dynamics algorithms. *Molecular Physics: An International Journal at the Interface Between Chemistry and Physics*, 65(6):1409–1419, 1988.

A.J Patterson, J Henrie-Olson, and R Brenner. Vasoregulation at the molecular level: a role for the beta1 subunit of the calcium-activated potassium (bk) channel. *Trends Cardiovasc Med*, 12(2):78–82, 2002.

L Pauling. The principles determining the structure of complex ionic crystals. *J Am Chem Soc*, 51(4):1010, 1929.

M Pavlov, P.E.M Siegbahn, and M Sandstrom. Hydration of beryllium, magnesium, calcium, and zinc ions using density functional theory. *The Journal of Physical Chemistry A*, 102(1):219, 1998.

W Pearson. Rapid and sensitive sequence comparison with fastp and fasta. *Methods Enzymol*, 183:63–98, 1990.

J Pei. Multiple protein sequence alignment. *Curr Opin Struct Biol*, 18(3):382–6, 2008.

G.J Perez, A.D Bonev, J.B Patlak, and M.T Nelson. Functional coupling of ryanodine receptors to kca channels in smooth muscle cells from rat cerebral arteries. *J Gen Physiol*, 113(2):229–38, 1999.

E Pettersen, T Goddard, C Huang, G Couch, D Greenblatt, E Meng, and T Ferrin. Ucsf chimera–a visualization system for exploratory research and analysis. *J Comput Chem*, 25(13):1605–12, 2004.

J.-P Piquemal, L Perera, G.A Cisneros, P Ren, L.G Pedersen, and T.A Darden. Towards accurate solvation dynamics of divalent cations in water using the polarizable amoeba force field: From energetics to structure. *The Journal of chemical physics*, 125(5):054511, 2006.

S Pluger, J Faulhaber, M Furstenau, M Lohn, R Waldschutz, M Gollasch, H Haller, F.C Luft, H Ehmke, and O Pongs. Mice with disrupted BK channel beta1 subunit gene feature abnormal ca(2+) spark/stoc coupling and elevated blood pressure. *Circ Res*, 87(11):E53–60, 2000.

J Ponder and D Case. Force fields for protein simulations. *Adv Protein Chem*, 66: 27–85, 2003. Force fields for protein simulations.

J.W Ponder, C Wu, P Ren, V.S Pande, J.D Chodera, M.J Schnieders, I Haque, D.L Mobley, D.S Lambrecht, R.A DiStasio, M Head-Gordon, G.N Clark, M.E Johnson, and T Head-Gordon. Current status of the amoeba polarizable force field. *J Phys Chem B*, 114(8):2549–64, 2010.

N Popovych, S Sun, R.H Ebright, and C.G Kalodimos. Dynamically driven protein allostery. *Nat Struct Mol Biol*, 13(9):831–8, 2006.

M Punta, L Forrest, H Bigelow, A Kernytsky, J Liu, and B Rost. Membrane protein prediction methods. *Methods*, 41(4):460–74, 2007.

A Rahman. Correlations in the motion of atoms in liquid argon. *Physical Review*, 136(2A):A405, 1964.

A Rahman and F Stillinger. Propagation of sound in water. a molecular-dynamics study. *Physical Review A*, 10(1):368, 1974.

K Ramanathan, T.H Michael, G.J Jiang, H Hiel, and P.A Fuchs. A molecular mechanism for electrical tuning of cochlear hair cells. *Science*, 283(5399):215–7, 1999.

M Rarey, B Kramer, and T Lengauer. Time-efficient docking of flexible ligands into active sites of proteins. *Proc Int Conf Intell Syst Mol Biol*, 3:300–8, 1995.

M Rarey, B Kramer, T Lengauer, and G Klebe. A fast flexible docking method using an incremental construction algorithm. *J Mol Biol*, 261(3):470–89, 1996.

P Ren and J Ponder. Polarizable atomic multipole water model for molecular mechanics simulation. *The Journal of Physical Chemistry B*, 107:5933–5947, 2003.

M Revington, Y Zhang, G.N Yip, A.V Kurochkin, and E.R Zuiderweg. Nmr investigations of allosteric processes in a two-domain thermus thermophilus hsp70 molecular chaperone. *J Mol Biol*, 349(1):163–83, 2005.

D Ritchie. Recent progress and future directions in protein-protein docking. *Curr Protein Pept Sci*, 9(1):1–15, 2008.

R Robitaille, E.M Adler, and M.P Charlton. Calcium channels and calcium-gated potassium channels at the frog neuromuscular junction. *J Physiol Paris*, 87(1): 15–24, 1993.

C Roessler, B Hall, W Anderson, W Ingram, S Roberts, W Montfort, and M Cordes. Transitive homology-guided structural studies lead to discovery of cro proteins with 40*Proc Natl Acad Sci U S A*, 105(7):2343–8, 2008.

C Rohl, C Strauss, D Chivian, and D Baker. Modeling structurally variable regions in homologous proteins with rosetta. *Proteins*, 55(3):656–677, 2004a.

C Rohl, C Strauss, K Misura, and D Baker. Protein structure prediction using Rosetta. *Methods Enzymol*, 383:66–93, 2004b.

B Rost and C Sander. Prediction of protein secondary structure at better than 70accuracy. *J Mol Biol*, 232(2):584–99, 1993.

B Rost, G Yachdav, and J Liu. The predictprotein server. *Nucleic Acids Res*, 32: 321–6, 2004.

B Roux. Computational studies of the gramicidin channel. *Acc Chem Res*, 35(6): 366–75, 2002.

P Sah and E.S Faber. Channels underlying neuronal calcium-activated potassium currents. *Prog Neurobiol*, 66(5):345–53, 2002.

A Sali and T.L Blundell. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*, 234(3):779–815, 1993.

A Sali, R Matsumoto, H McNeil, M Karplus, and R Stevens. Three-dimensional models of four mouse mast cell chymases. identification of proteoglycan binding regions and protease-specific antigenic epitopes. *J Biol Chem*, 268(12):9023–34, 1993.

A Saxena, D Wong, K Diraviyam, and D Sept. The basic concepts of molecular modeling. In *Methods in Enzymology*, volume 467, pages 307 – 334. Academic Press, 2009.

R Sayle and E Milner-White. Rasmol: biomolecular graphics for all. *Trends Biochem Sci*, 20(9):374, 1995.

H Schafer, L.J Smith, A.E Mark, and W.F van Gunsteren. Entropy calculations on the molten globule state of a protein: side-chain entropies of alpha-lactalbumin. *Proteins*, 46(2):215–24, 2002.

M Schreiber and L Salkoff. A novel calcium-sensing domain in the BK channel. *Biophys J*, 73(3):1355–63, 1997.

M Schreiber, A Wei, A Yuan, J Gaut, M Saito, and L Salkoff. Slo3, a novel ph-sensitive K+ channel from mammalian spermatocytes. *J Biol Chem*, 273(6):3509–16, 1998.

M Schreiber, A Yuan, and L Salkoff. Transplantable sites confer calcium sensitivity to BK channels. *Nat Neurosci*, 2(5):416–21, 1999.

Schrödinger, LLC. The PyMOL molecular graphics system, version 1.3r1. August 2010.

O Schueler-Furman, C Wang, and D Baker. Progress in protein-protein docking: atomic resolution predictions in the capri experiment using Rosettadock with an improved treatment of side-chain flexibility. *Proteins*, 60(2):187–94, 2005a.

O Schueler-Furman, C Wang, P Bradley, K Misura, and D Baker. Progress in modeling of protein structures and interactions. *Science*, 310(5748):638–42, 2005b.

T Schulz-Gasch and M Stahl. Binding site characteristics in structure-based virtual screening: evaluation of current docking tools. *J Mol Model*, 9(1):47–57, 2003.

T Schwede, J Kopp, N Guex, and M Peitsch. Swiss-model: An automated protein homology-modeling server. *Nucleic Acids Res*, 31(13):3381–5, 2003.

K.Z Shen, A Lagrutta, N.W Davies, N.B Standen, J.P Adelman, and R.A North. Tetraethylammonium block of slowpoke calcium-activated potassium channels expressed in xenopus oocytes: evidence for tetrameric channel formation. *Pflugers Arch*, 426(5):440–5, 1994.

J Shi, T Blundell, and K Mizuguchi. Fugue: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J Mol Biol*, 310(1):243–57, 2001.

J Shi, G Krishnamoorthy, Y Yang, L Hu, N Chaturvedi, D Harilal, J Qin, and J Cui. Mechanism of magnesium activation of calcium-activated potassium channels. *Nature*, 418(6900):876–80, 2002.

B Shoichet. Virtual screening of chemical libraries. *Nature*, 432(7019):862–5, 2004.

B Shoichet and I Kuntz. Protein docking and complementarity. *J Mol Biol*, 221(1): 327–46, 1991.

B Shoichet and I Kuntz. Matching chemistry and shape in molecular docking. *Protein Eng*, 6(7):723–32, 1993.

D Shortle, K Simons, and D Baker. Clustering of low-energy conformations near the native structures of small proteins. *Proc Natl Acad Sci U S A*, 95(19):11158–62, 1998.

K Simons, C Kooperberg, E Huang, and D Baker. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions. *J Mol Biol*, 268(1):209–25, 1997.

K Simons, R Bonneau, I Ruczinski, and D Baker. Ab initio protein structure prediction of casp iii targets using rosetta. *Proteins*, Suppl 3:171–6, 1999a.

K Simons, I Ruczinski, C Kooperberg, B Fox, C Bystroff, and D Baker. Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. *Proteins*, 34(1):82–95, 1999b.

S Sousa, P Fernandes, and M Ramos. Protein-ligand docking: current status and future challenges. *Proteins*, 65(1):15–26, 2006.

M Stahl and T Schulz-Gasch. Practical database screening with docking tools. *Ernst Schering Res Found Workshop*, 42:127–51, 2003.

N Standen. Cardiovascular biology: Tuning channels for blood pressure. *Nature*, 407 (6806):845–848, Oct 2000. 10.1038/35038185.

A M Stock, E Martinez-Hackert, B F Rasmussen, A H West, J B Stock, D Ringe, and G A Petsko. Structure of the mg(2+)-bound form of chey and mechanism of phosphoryl transfer in bacterial chemotaxis. *Biochemistry*, 32(49):13375–80, Dec 1993.

J.B Stock, A.M Stock, and J.M Mottonen. Signal transduction in bacteria. *Nature*, 344(6265):395–400, 1990.

M.J Stone. Nmr relaxation studies of the role of conformational entropy in protein stability and ligand binding. *Acc Chem Res*, 34(5):379–88, 2001.

N.C Strynadka and M.N James. Crystal structures of the helix-loop-helix calcium-binding proteins. *Annu Rev Biochem*, 58:951–98, 1989.

S Subramaniam. The biology workbench–a seamless database and analysis environment for the biologist. *Proteins*, 32(1):1–2, 1998.

L Svensson, Eva Thulin, and Sture Forsén. Proline cis-trans isomers in Calbindin D9k observed by x-ray crystallography. *Journal of Molecular Biology*, 223(3):601–606, Feb 1992.

J.F Swain, G Dinler, R Sivendran, D.L Montgomery, M Stotz, and L.M Gierasch. Hsp70 chaperone ligands control domain association via an allosteric mechanism mediated by the interdomain linker. *Mol Cell*, 26(1):27–39, 2007.

D.M Szebenyi and K Moffat. The refined structure of vitamin d-dependent calcium-binding protein from bovine intestine. molecular details, ion binding, and implications for the structure of other calcium-binding proteins. *J Biol Chem*, 261(19): 8761–77, 1986.

J.D Thompson, D.G Higgins, and T.J Gibson. Clustal w: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, 22(22):4673–80, 1994.

D Tieleman and H Berendsen. Molecular dynamics simulations of a fully hydrated dipalmitoylphosphatidylcholine bilayer with different macroscopic boundary conditions and parameters. *The Journal of chemical physics*, 105(11):4871, 1996.

D Tieleman, S Marrink, and H Berendsen. A computer perspective of membranes: molecular dynamics studies of lipid bilayer systems. *Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes*, 1331(3):235, 1997.

D.P Tieleman and H.J Berendsen. A molecular dynamics study of the pores formed by escherichia coli ompf porin in a fully hydrated palmitoyloleoylphosphatidylcholine bilayer. *Biophys J*, 74(6):2786–801, 1998.

R Tiwari, K Mahasenan, R Pavlovicz, C Li, and W Tjarks. Carborane clusters in computational drug design: a comparative docking evaluation using autodock, flexx, glide, and surflex. *J Chem Inf Model*, 49(6):1581–9, 2009.

A Toker and R Macnab. Distinct regions of bacterial flagellar switch protein flim interact with flig, flin and chey. *Journal of Molecular Biology*, 273(3):623–634, Oct 1997.

A Tongraar and B.M Rode. Structural arrangement and dynamics of the hydrated Mg2+: An ab initio qm/mm molecular dynamics simulation. *Chem Phys Lett*, 409 (4-6):304, 2005.

L Toro, M Wallner, P Meera, and Y Tanaka. Maxi-kca, a unique member of the voltage-gated k channel superfamily. *News Physiol Sci*, 13(3):112–117, 1998.

V Tsui and D.A Case. Theory and applications of the generalized born solvation model in macromolecular simulations. *Biopolymers*, 56(4):275–91, 2000.

S Vajda and D Kozakov. Convergence and combination of methods in protein-protein docking. *Curr Opin Struct Biol*, 19(2):164–70, 2009.

J Voigt, B Bienfait, S Wang, and M Nicklaus. Comparison of the nci open database with seven large chemical structural databases. *J Chem Inf Comput Sci*, 41(3): 702–12, 2001.

K Volz and P Matsumura. Crystal structure of escherichia coli chey refined at 1.7-a resolution. *J Biol Chem*, 266(23):15511–9, 1991.

G Vriend. What if: a molecular modeling and drug design program. *J Mol Graph*, 8 (1):52–6, 29, 1990.

G Wadhams and J Armitage. Making sense of it all: bacterial chemotaxis. *Nat Rev Mol Cell Biol*, 5(12):1024–1037, 2004. 10.1038/nrm1524.

B Wang, B S Rothberg, and R Brenner. Mechanism of increased BK channel activation from a channel mutation that causes epilepsy. *J Gen Physiol*, 133(3):283–94, Mar 2009.

C Wang, O Schueler-Furman, I Andre, N London, S Fleishman, P Bradley, B Qian, and D Baker. Rosettadock in capri rounds 6-12. *Proteins*, 69(4):758–63, 2007.

L Wang and F Sigworth. Structure of the BK potassium channel in a lipid membrane from electron cryomicroscopy. *Nature*, 461(7261):292–295, Sep 2009. 10.1038/nature08291.

Z.W Wang, O Saifee, M.L Nonet, and L Salkoff. Slo-1 potassium channels control quantal content of neurotransmitter release at the c. elegans neuromuscular junction. *Neuron*, 32(5):867–81, 2001.

G Warren, C Andrews, A Capelli, B Clarke, J LaLonde, M Lambert, M Lindvall, N Nevins, S Semus, S Senger, G Tedesco, I Wall, J Woolven, C Peishoff, and M Head. A critical assessment of docking programs and scoring functions. *J Med Chem*, 49(20):5912–31, 2006.

A Waterhouse, J Procter, D Martin, M Clamp, and G Barton. Jalview version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9): 1189–91, 2009.

M Wiederstein and M Sippl. Prosa-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res*, 35:407–10, 2007.

C Withers-Martinez, F Carriere, R Verger, D Bourgeois, and C Cambillau. A pancreatic lipase with a phospholipase a1 activity: crystal structure of a chimeric pancreatic lipase-related protein 2 from guinea pig. *Structure*, 4(11):1363–74, 1996.

D W.Mount. *Bioinformatics:Sequence and Genome Analysis*. 2001. Bioinformatics:Sequence and Genome Analysis.

K Wolf-Maier, R.S Cooper, H Kramer, J.R Banegas, S Giampaoli, M.R Joffres, N Poulter, P Primatesta, B Stegmayr, and M Thamm. Hypertension treatment and control in five european countries, canada, and the united states. *Hypertension*, 43(1):10–7, 2004.

Y Wu, Y Yang, S Ye, and Y Jiang. Structure of the gating ring from the human large-conductance Ca2+-gated K+ channel. *Nature*, 466(7304):393–397, Jul 2010.

X.M Xia, X Zeng, and C.J Lingle. Multiple regulatory sites in large-conductance calcium-activated potassium channels. *Nature*, 418(6900):880–4, 2002.

W Xu, Y Liu, S Wang, T McDonald, J.E Van Eyk, A Sidor, and B O'Rourke. Cytoprotective role of Ca2+- activated K+ channels in the cardiac inner mitochondrial membrane. *Science*, 298(5595):1029–33, 2002.

T Yamaguchi, S Hayashi, and H Ohtaki. X-ray diffraction study of calcium(II) chloride hydrate melts: CaCl2.RH2O (R = 4.0, 5.6, 6.0, and 8.6). *Inorganic Chemistry*, 28(12):2434, 1989.

H Yang, L Hu, J Shi, and J Cui. Tuning magnesium sensitivity of BK channels by mutations. *Biophys J*, 91(8):2892–900, 2006.

H Yang, L Hu, J Shi, K Delaloye, F.T Horrigan, and J Cui. Mg^{2+} mediates interaction between the voltage sensor and cytosolic domain to activate BK channels. *Proc Natl Acad Sci U S A*, 104(46):18270–5, 2007.

J Yang, G Krishnamoorthy, A Saxena, G Zhang, J Shi, H Yang, K Delaloye, D Sept, and J Cui. An epilepsy/dyskinesia-associated mutation enhances BK channel activation by potentiating Ca^{2+} sensing. *Neuron*, 66(6):871–83, 2010.

V Yarov-Yarovoy, D Baker, and W.A Catterall. Voltage sensor conformations in the open and closed states in rosetta structural models of K^+ channels. *Proc Natl Acad Sci U S A*, 103(19):7292–7, 2006.

M.A Young, S Gonfloni, G Superti-Furga, B Roux, and J Kuriyan. Dynamic coupling between the sh2 and sh3 domains of c-src and hck underlies their inactivation by c-terminal tyrosine phosphorylation. *Cell*, 105(1):115–26, 2001.

H Yu, T Whitfield, E Harder, G Lamoureux, I Vorobyov, V Anisimov, A MacKerell, and B Roux. Simulating monovalent and divalent ions in aqueous solution using a drude polarizable force field. *Journal of Chemical Theory and Computation*, 6(3): 774–786, Mar 2010. doi: 10.1021/ct900576a.

A Yuan, M Dourado, A Butler, N Walton, A Wei, and L Salkoff. Slo-2, a K+ channel with an unusual cl- dependence. *Nat Neurosci*, 3(8):771–9, 2000.

A Yuan, C.M Santi, A Wei, Z.W Wang, K Pollak, M Nonet, L Kaczmarek, C.M Crowder, and L Salkoff. The sodium-activated potassium channel is encoded by a member of the slo gene family. *Neuron*, 37(5):765–73, 2003.

P Yuan, M Leonetti, A Pico, Y Hsiung, and R MacKinnon. Structure of the human BK channel Ca2+-activation apparatus at 3.0 a resolution. *Science*, 329(5988): 182–186, Jul 2010.

T Yusifov, N Savalli, C.S Gandhi, M Ottolia, and R Olcese. The rck2 domain of the human BKca channel is a calcium sensor. *Proc Natl Acad Sci U S A*, 105(1): 376–81, 2008.

E Zdobnov and R Apweiler. Interproscan–an integration platform for the signature-recognition methods in interpro. *Bioinformatics*, 17(9):847–8, 2001.

G Zhang, Sheng-You Huang, J Yang, J Shi, X Yang, A Moller, X Zou, and J Cui. Ion sensing in the rck1 domain of BK channels. *Proceedings of the National Academy of Sciences*, 107(43):18700–18705, Oct 2010.

L. Zhang, X. Li, R. Zhou, and G. Xing. Possible role of potassium channel, big k in etiology of schizophrenia. *Medical hypotheses*, 67(1):41–43, 2006.

Y Zhang. I-tasser server for protein 3d structure prediction. *BMC Bioinformatics*, 9: 40, 2008.

Z Zhang, A Schaffer, W Miller, T Madden, D Lipman, E Koonin, and S Altschul. Protein sequence similarity searches using patterns as seeds. *Nucleic Acids Res*, 26 (17):3986–90, 1998.

X Zhu, J Rebello, P Matsumura, and K Volz. Crystal structures of chey mutants Y106W and T87I/Y106W. chey activation correlates with movement of residue 106. *J Biol Chem*, 272(8):5000–6, 1997.

R ZhuGe, K.E Fogarty, R.A Tuft, L.M Lifshitz, K Sayar, and J.V Walsh. Dynamics of signaling between ca(2+) sparks and ca(2+)- activated k(+) channels studied with a novel image-based method for direct intracellular measurement of ryanodine receptor ca(2+) current. *J Gen Physiol*, 116(6):845–64, 2000.

L Zidek, M.V Novotny, and M.J Stone. Increased protein backbone conformational entropy upon hydrophobic ligand binding. *Nat Struct Biol*, 6(12):1118–21, 1999.

V Zoete, A Grosdidier, and O Michielin. Docking, virtual high throughput screening and in silico fragment-based drug design. *J Cell Mol Med*, 13(2):238–48, 2009.

Curriculum vitae

Akansha Saxena
akanshasaxena@gmail.com

Degrees

B.Tech. May 2003
Biotechnology and Biochemical Engineering
Indian Institute of Technology, Kharagpur, India

M.Tech. May 2004
Biotechnology and Biochemical Engineering
Indian Institute of Technology, Kharagpur, India

M.S. December 2007
Biomedical Engineering
Washington University in St Louis, St Louis, MO, USA

Ph.D. December 2010
Biomedical Engineering
Washington University in St Louis, St Louis, MO, USA

Employment

Visiting Research Investigator, 7/2009 - 12/2010
University of Michigan, Ann Arbor, MI, USA

Research Assistant, 5/2003 - 5/2004
Indian Institute of Technology, Kharagpur, India

Research Assistant, Summers, 5/2002 - 7/2002
Center for Biochemical Technology, Delhi University, India

Research Assistant, Summers, 6/2001 - 7/2001
Bio-Engineering and Research Centre
Indian Institute of Technology, Delhi, India

Professional Societies

Biophysical Society
American Chemical Society

Peer-reviewed publications

Saxena, A, Wong, D., Diraviyam, K., and Sept, D., The Basic Concepts of Molecular Modeling, Methods Enzymol. 2009; 467:307- 34

Yang J.*, Krishnamoorthy, G*., **Saxena, A***., Zhang, G., Shi, H., Yang, J., Delaloyle, K., Sept, D. and Cui, J. An Epilepsy/Dyskinesia-Associated Mutation Enhances BK Channel Activation by Potentiating Ca2+ Sensing, Neuron, Volume 66, Issue 6, 871-883, 24 June 2010 (*co-first authors)

Saxena, A. and Sept, D. Structural and dynamical effect of the D369G mutation on the RCK1 domain of BK channel. (In preparation)

Saxena, A. and Sept, D. Development of multisite cation models to improve the free energy calculations in molecular dynamics simulations. (In preparation)

Invited Presentations

"The epilepsy mutation potentiates the Ca^{2+} activation mechanism in BK_{Ca} channels", Biochemistry and Molecular Biophysics Program Retreat, St Louis, MO, Oct, 2008

"Prediction of a high affinity Ca^{2+} binding site in the BK_{Ca} channel structure", 2008 Biomedical Engineering Society Annual Fall Meeting, St Louis, MO, Oct, 2008

"Probing the role of epilepsy mutation in BK_{Ca} channel activation through computer simulations", CBB Seminar, Gerstein Lab, Yale University, New Haven, CT, July 2010

"Structure-Function studies on BK_{Ca} channels and Multisite Cation Model", Garcia Lab Seminar, Rensselaer Polytechnic Institute, Troy, NY, December 2010

December 2010